AUTODESK UNIVERSITY 2014

A Hardware Wonk's Guide to Specifying the Best Building Information Modeling and 3D Computing Workstations, 2014 Edition

Matt Stachoni – BIM / IT Manager, Erdy McHenry Architecture LLC

CM6739 Working with today's Building Information Modeling (BIM) tools presents a special challenge to your IT infrastructure. As you wrestle with the computational demands of the Revit software platform—as well as with high-end graphics in 3ds Max Design, Showcase, and Navisworks Manage—you need the right knowledge to make sound investments in your workstation and server hardware. Get inside the mind of a certified (some would say certifiable) hardware geek and understand the variables to consider when purchasing hardware to support the demands of these BIM and 3D products from Autodesk, Inc. Fully updated for 2014, this class gives you the scoop on the latest advancements in workstation gear, including processors, motherboards, memory, and graphics cards. This year we also focus on the IT closet, specifying the right server gear, and high-end storage options.

Learning Objectives

At the end of this class, you will be able to:

- Discover the current state of the art and "sweet spots" in processors, memory, storage, and graphics
- Optimize your hardware resources for BIM modeling, visualization, and construction coordination
- Understand what is required in the IT room for hosting Autodesk back-end services like Revit Server application and Vault software
- Answer the question, "Should I build or should I buy?"

About the Speaker

Matt is the BIM and IT Manager for Erdy McHenry Architecture LLC, an architectural design firm in Philadelphia, Pennsylvania. He is responsible for the management, training, and support of the firm's digital design and BIM efforts. He continuously conducts R&D on new application methodologies, software and hardware tools, and design platforms, applying technology to theory and professional practice. He specifies, procures, and implements IT technology of all kinds to maximize the intellectual capital spent on projects.

Prior to joining Erdy McHenry, Matt was a senior BIM implementation and IT technical specialist for CADapult Ltd., an Authorized Autodesk Silver reseller servicing the Mid-Atlantic region. There, he provided training for AEC customers, focused primarily on implementing BIM on the Revit platform, Navisworks, and related applications. Matt also provided specialized BIM support services for the construction industry, such as construction modeling, shop drawing production, and project BIM coordination.

Matt has been using Autodesk® software since 1987 and has over 20 years' experience as a CAD and IT Manager for several A/E firms in Delaware, Pennsylvania, and Boston, Massachusetts. He is a contributing writer for AUGIWorld Magazine and this is his 11th year speaking at Autodesk University.

Email: matt@stachoni.com
mstachoni@em-arc.com

Twitter: @MattStachoni

Section I: Introduction

Building out a new BIM / 3D workstation specifically tuned for Autodesk's Building Design Suite can quickly become confusing with all of the choices you have. Making educated guesses as to where you should spend your money - and where you should not - requires time to research through product reviews, online forums, and working with salespeople who don't understand what you do on a daily basis. Advancements in CPUs, GPUs, and storage can test your notions of what is important and what is not.

Computing hardware had long ago met the relatively low demands of 2D CAD, but data-rich 3D BIM and visualization still presents a challenge. New Revit and BIM users will quickly learn that the old CAD-centric rules for specifying workstations no longer apply. You are not working with many small, sub-MB files. BIM applications do not fire up on a dime. Project assets can easily exceed 1GB as you create rich datasets with comprehensive design intent and construction BIM models, high resolution 3D renderings, animations, Photoshop files, and so on. Simply put, the extensive content you create using one or all of the applications in the Building Design Suite requires the most powerful workstations you can afford.

Additionally, each of the tools in the Suite get more complex as their capability improves with each release. Iterating through adaptive components in Revit, or using the newer 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 Design®, Navisworks®, and Showcase®. We focus on three key areas.

We want to answer these fundamental questions:

- What aspects of your system hardware does each application in the Building Design Suite stress?
- What are the appropriate choices in processors today, and which are not?
- How much system RAM is appropriate? Where does it make a difference?
- What's the difference between a workstation graphics card and a "gaming" card?
- Are solid state drives (SSDs) worth the extra cost? What size should I go for?
- What's new in mobile workstations?
- I have a screwdriver and I know how to use it. Do I build my own machine or do I buy a complete system from a vendor?

To do this we will look at computing subsystems in detail, and review the important technical aspects you should consider when choosing a particular component:

- Central Processing Units (CPUs)
- Chipsets and motherboard features
- System memory (RAM)
- Graphics processors (GPUs)
- Storage
- Peripherals – Displays, mice, and keyboards

Disclaimer

In this class I will often make references and tacit recommendations for specific system components. This is my opinion, largely coming 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. You are encouraged 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.

Industry Pressures and Key Trends

The AEC design industry has quickly migrated from traditional 2D, CAD-centric applications and methodologies to intelligent, model-based ones. 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 subsystem is challenged by a particular task.

Similarly for PC technologies there exist several key areas which are shaping the future of today's high-end computing: Maximizing Performance per Watt (PPW), recognizing the importance of multithreading and multiprocessing performance, leveraging GPU-accelerated computing, and increased implementation of cloud computing. Taken together these technologies allow us to scale up, down, and out.

Performance per Watt

It may come as a surprise to learn that, for any single component, the increase of raw performance in this year's model over last year's is by itself is no longer of primary importance for manufacturers. Instead, increasing the efficiency of components is a paramount design criteria, which essentially maximizes Performance per Watt (PPW).

This is largely due to the mass movement in CPUs, graphics, and storage towards smaller and more mobile technologies. Cell phones, tablets, laptops, and mobile workstations are more appealing than desktop computers but have stringent energy consumption constraints which limit performance bandwidth. Increasing PPW allows higher performance to be stuffed into smaller and more mobile platforms.

This has two side effects. Mobile technologies are making their way into desktop components, allowing for CPUs and GPUs that are more energy efficient, run cooler, and are quiet. This means you can have more of them in a single workstation.

The other side effect is that complex BIM applications can be extended from the desktop to more mobile platforms, such as performing 3D modeling using a small laptop during design meetings, running clash detection on the construction site using tablets, or using drone-mounted cameras to turn HD imagery into fully realized 3D models.

Parallel Processing

The key problems associated with BIM and 3D visualization, such as energy modeling and high-end visualization, are often too big for a single processor or computer system to handle efficiently. However, many of these problems are highly parallel in nature, where separate calculations are carried out simultaneously and independently. Large tasks can often be neatly broken down into smaller ones that don't rely on each other to finish before being worked on. Accordingly, these kinds of workloads can be distributed to multiple processors or even out to multiple physical computers, each of which can chew on that particular problem and return results that can be aggregated later.

In particular, 3D photorealistic visualization 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 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.

This processing of millions of rays can readily be broken down into chunks of smaller tasks that can be handled independently. Accordingly, the more CPUs you can throw at a rendering task the faster it will finish. In fact, you can pipe the task out to multiple physical machines to work on the problem.

Discreet and Autodesk recognized the benefits of parallel processing early on in 3ds Max, and promoted the idea of disseminating a rendering process across separate machines using Backburner. You can easily create a rendering farm where one machine sends a rendering job 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. What would take a single PC hours can be created in a fraction of the time with enough machines.

Multiprocessing and Multithreading

Just running an operating system and separate applications is, in many ways, a parallel problem as well. Even without running a formal application, a modern OS has many smaller processes running at the same time, such as the security subsystem, anti-virus protection, network connectivity, etc. Each of your applications may run one or more separate processes on top of that, and processes themselves can spin off separate threads of execution.

All modern processors and operating systems fully support both **multiprocessing,** the ability to push separate processes to multiple CPUs 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 on a single CPU. 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.

We've all made the mass migration to multi-core computing, even down to our tablets and phones. Today you can maximize both, and outfit a high-end workstation to have multiple physical CPUs, each with multiple cores, which substantially increases a single machine's performance.

The Road to GPU Accelerated Computing

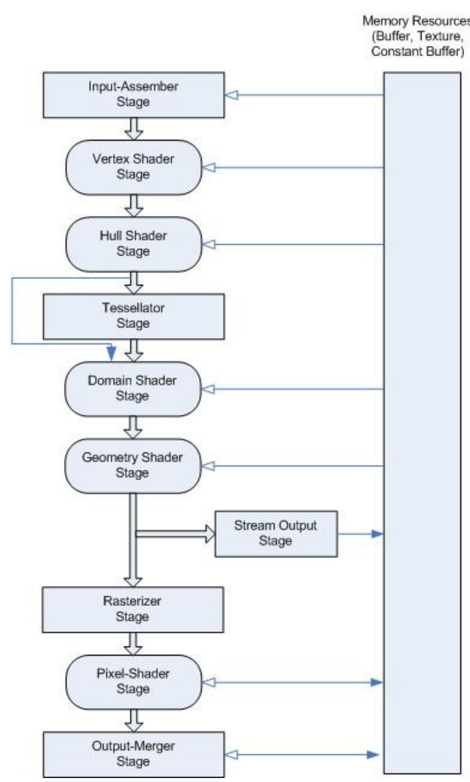
Multiprocessing is not limited to CPUs any longer. Recognizing the parallel nature of many graphics tasks, GPU designers at ATI and NVIDIA have created GPU architectures for their graphics cards that are massively multiprocessing in nature. As a result we can now offload compute-intensive portions of a problem to the GPU and free the CPU up to run other code. And those tasks do not have to be graphics related, but could focus on things like modeling storm weather patterns, acoustics, protein folding, etc.

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 therefore designed such that more transistors are devoted to data processing rather than data caching and flow control.

A CPU consists of relatively few cores – from 2 to 8 in most systems - 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 smaller, highly efficient cores designed to execute many concurrent threads more slowly. These are often referred to as Stream Processors. Indeed, it is by increasing Performance per Watt that the GPU can cram so many cores into a single die.

It wasn't always like this. Back in the day, 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 their functionality to dedicated logic in the GPU that was designed to support them in a hard-coded fashion.

A traditional graphics data pipeline is really a rasterization pipeline. It is composed of a series of steps used to create a 2D raster representation of 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.



In games, the GPU needs to do this as many times a second as possible to maintain smoothness of play. Accuracy and photorealism are sacrificed for speed. Games don't render a car that reflects the street correctly because they can't. But they can still display highly complex graphics and effects. How?

Today's GPUs have a programmable graphics pipeline which can be manipulated through small programs called Shaders, which are specialized programs that make complex effects happen in real time. OpenGL and Direct3D (DirectX) 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 throughout. The DirectX API, released with each version of Windows, regularly defines new shader models which increase programming model flexibilities and capabilities.

However, traditional ray-tracing rendering engines such as NVIDIA's mental ray did not use the computational power of the GPU to handle the ray-tracing algorithms. Instead, rendering was almost entirely a CPU-bound operation, in that it doesn't rely much (or at all) on the graphics card to produce the final image. Designed to pump many frames to the screen per second, GPUs were not meant to do the kind of detailed ray-tracing calculation work on a single static image in real time.

That is rapidly changing as most of the GPU 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) 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.

Today, the most compelling area GPU Compute comes into play for Building Design Suite users is the iRay rendering engine in 3ds Max Design. We'll discuss this in more depth in the section on graphics. However, in the future I would not be surprised to see GPU compute technologies to be exploited for other uses across BIM applications.

Virtualization

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 almost every business today in some fashion. Most companies employ them in the server closet, hosting multiple VMs on a single server-class box. This allows a company employ fewer physical machines to host file storage servers, Microsoft Exchange servers, SQL database servers, application servers, web servers, and others. For design firms, 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.

VMs also allows IT a lot of flexibility in how these servers are apportioned across available hardware and allows for better serviceability. 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 installed on a normal server.

IT may 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 an XP VM and run it under that.

Today's virtualization extends to the workstation as well. Companies are building out their own on premise clouds in their data 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 2015 releases).

The primary limitation for deploying VMs for use for high-end applications like Revit, Navisworks, and 3ds Max has been in the graphics department. Simply put, 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 and provide them and all of their power to VMs hosted on that box.

The Cloud Effect

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 but the per process costs drop. This economy of scale has 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, or "Cloud" computing services. The Cloud is quickly migrating from the low hanging fruit of simple storage-anywhere-anytime mechanism (e.g., Dropbox, Box.net), to massive remote access capabilities to fast machines which will soon become on-demand, essentially limitless, very cheap computing horsepower.

As such, 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 360's 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.

Unfortunately as of this writing it only works within Revit, AutoCAD, and Navisworks, and does not work with 3ds Max, Maya, or other 3D applications such as SketchUp or Rhino. For these applications there are hundreds of 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.

Even general-purpose cloud-processing providers such as Amazon's EC2 service provide 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. 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 is here: http://www.migenius.com/products/NVIDIA-iray/iray-benchmarks and a tutorial on how to set up an EC2 account is here: http://area.autodesk.com/blogs/cory/setting-up-an-amazon-ec2-render-farm-with-backburner

We can see where the future is leading, that is, to “thin” desktop clients with just enough computing horsepower 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 will at some point no longer need to 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 for CPUs, RAM, storage, etc., the pricing differences between any two similar components of different capacities or speeds has shrunk, making the higher end option a more logical buy. For example, a high quality 1TB drive is about $70, a 2TB drive is about $130, and a 3TB drive is about $145 more than that, so you get 3x the storage for about 2x the price. Get the higher capacity drive and you likely not worry about upgrading for far longer.

For system memory, conventional wisdom once decreed 8GB as a starting point for BIM applications, but not today. This first meant going with 4x2GB 240-pin DDR3 memory modules, as 4GB modules were expensive at the time. Today, a 2GB module is about $35 ($17.50/GB), and 4GB modules have dropped to about $37 ($9.25/GB), making it less expensive to outfit the system with 2x4GB modules. However, 8GB modules have now dropped to about $70, or only $8.75/GB.

Thus, for a modest additional investment it makes more sense to install 16GB as 2x8GB modules as a base point for any new BIM system. Most desktop motherboards have 4 memory slots, so you can max out the system with 32GB (4x8GB) and not worry about RAM upgrades at all. Note that mainstream desktop CPUs like the Core i7-4790 (discussed later) won’t see more than 32GB of RAM anyway.

In both of these cases 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 500GB drive is more than adequate to house the OS, applications, and a user’s profile data.

Processors have a different story. CPU pricing is based upon capability and popularity, but price curves are anything but linear. A 3.2GHz CPU might be $220 and a 3.4GHz incrementally higher at $250, but a 3.5GHz CPU 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 (graphics processing unit) 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 with model variations in base clock, onboard memory, or number of internal GPU cores present or activated. Both companies issue reference boards that card manufacturers use to build their offerings. Thus, pricing between different manufacturer’s cards with 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.

Shrinking prices for components that are good enough for the mainstream can skew the perception of what a machine should cost for heavy-duty database and graphics processing in Revit, Navisworks and other BIM applications. 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.

Building Design Suite Application Demands

Within each workstation there are four primary component that affect overall performance: the processor (CPU), system memory (RAM), the graphics card (GPU), and the storage subsystem. Each application within the Building Design Suite will stress these four components in different ways and to different extremes. Given the current state of hardware, today's typical entry-level workstation may perform well in most of the apps within the Suite, but not all, due to specific deficiencies in one or more system components. 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.

	CPU Speed / Multithreading	System Ram - Amount / Speed	Graphics Card GPU Capabilities	Graphics Card Memory Size	Hard Drive Speed
Revit	10 / 9	10 / 7	5	5	10
3ds Max Design	10 / 10	9 / 7	7 / 5 /10 (Nitrous / mr / iRay)	6 / 10 (mr / iRay)	10
Navisworks Simulate **Navisworks Manage**	8 / 7	7 / 6	7	5	8
Showcase	9 / 8	8 / 6	9	5	9
AutoCAD (2D & 3D)	6 / 6	5 / 5	5	5	6
AutoCAD Architecture **AutoCAD MEP**	8 / 6	7 / 5	5	5	6
ReCap Studio / Pro	10 / 10	9 / 5	8	7	10

Let's define an "entry-level workstation" to include the following base level components:

- CPU: Intel Third-Generation (Ivy Bridge) Quad-Core Core i5-3570K @ 3.4GHz, 6MB L3 cache
- System RAM: 8GB DDR3-1333
- Graphics Card: ATI Radeon 5750 1GB PCIe / NVIDIA GT 310 (c. 2010)
- Storage: 500GB 7200 RPM hard disk

The entry-level workstation defined above will perform adequately well in these applications up to a rating of about 7. For example, you can see that such a system will be enough for AutoCAD and its verticals, but would want some tweaking to run higher-order apps like Navisworks Manage, and is really inappropriate for Revit or 3ds Max Design. Not that those applications will not run in such a baseline system; but rather, that system is not optimized for those applications. Later we will be talking about specific components and how each affects our applications.

For application / component ratings over 6, you need to carefully evaluate your needs in each application and specify more capable parts. 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.

Application Notes: Revit

Autodesk Revit is rather unique in that the platform stresses every major component in a computer in ways that typical desktop applications do not. Users of the Building Design Suite will spend more hours per day in Revit than most other applications, so tuning your workstation specifically for Revit is a smart choice.

Because of the size and complexity of most BIM projects, it requires the fastest CPU, the most RAM, and the fastest storage system available. On the graphics side, Revit has rather mundane graphics demands. We've found that most can get by with relatively medium-powered cards, even on large projects.

Revit is, at its heart, a database management 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 at the very high end of the spectrum, where CPUs geared for database servers 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 of this work is computationally expensive. There are no shortcuts to be had; it has to do everything by the numbers to ensure model fidelity. It 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. When you have 8+ people doing this, things can and do get slow.

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 latest microarchitecture platform, the most cores, the fastest core clock speed, and the most L2 cache available. We will discuss these specific options in the Processor section of this handout.

Revit supports multi-threading in certain operations:

- Vector printing
- 2D Vector Export such as DWG and DWF
- Rendering
- Wall Join representation in plan and section views
- Loading elements into memory reduces view open times when elements are initially displayed
- Parallel computation of silhouette edges when navigating perspective 3D views
- Translation of high level graphical representation of model elements and annotations into display lists optimized for a given graphics card. Engaged when opening views or changing view properties
- File Open and Save
- Point Cloud Data Display

Autodesk will continue to exploit these kinds of improvements in other areas in future releases.

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, we have found Revit performs adequately well on most projects under relatively mainstream (between $100 and $300) graphics cards.

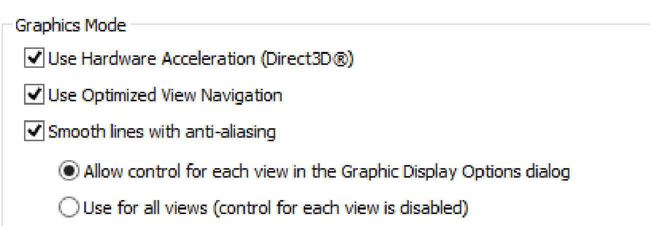
This is mostly because Revit views typically contain only a subset of the total project geometry. Most views are 2D, so the most Revit has to really do is perform lots of Hide operations. Even in 3D views, one typically filters out and limit the amount of 3D data which enables the system to respond quickly enough for most GPUs can handle with aplomb.

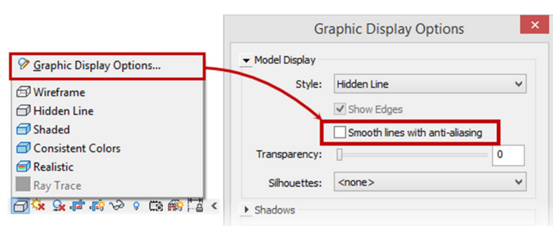
But as we use Revit as our primary 3D design and modeling application, the graphics card gets a real workout as we demand the ability to spin around our building quickly, usually in a shaded view. Toss in material appearances in Realistic view mode, new sketchy lines in 2015, anti-aliasing, ambient shadows, 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 got the ability to limit smoothing / anti-aliasing operations on a per-view setting using the Graphics Display Options dialog. Anti-aliasing 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 improves performance in the Ray Trace interactive rendering visual style, providing faster, higher quality rendering with improved color accuracy and shadows with all backgrounds. In other views, 2015 improves drawing performance such that many elements are drawn simultaneously in larger batches using fewer drawing calls. 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.

While Revit does want a decent graphics card foundation for upper order operations, it is completely agnostic about specific video card makes or models. All cards manufactured over the past four years will support Revit 2015's minimum requirement of DirectX 11 / Shader Model 3 under Windows 7 64-bit, which will allow for all viewport display modes, adaptive degradation, ambient occlusion effects, and so on. The general rule that the faster (and more expensive) the card is, the better it will be for Revit certainly applies, but only to a point with mainstream models. You probably would not see any real differences between mainstream and high end cards until you work with very large (over 800MB) models. You will most likely see zero difference between a $300 GeForce GTX and a $5,000 Quadro K6000.

Storage: Now look at the files you are creating - they are huge compared to traditional CAD files and represent a bottleneck in opening and saving projects. 60MB Revit files are typical minimums for smaller projects under 75,000 square feet, with 100MB being more common. MEP models with typically start around 60-80MB for complete projects and go up from there. On larger, more complex models (particularly those used for construction), expect file sizes to grow well over 300MB. 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 1Gbps networks these days), these operations take a toll on your local storage subsystem.

Finally, don't forget that Revit itself is a large program and takes a while just to fire up, so you need 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 directly tied to implementing efficient Best Practices in your company. An inefficient 200MB model will perform much worse than a very efficient 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. The Model Performance Technical Note 2014 has been updated from the previous version (2010) and is an invaluable resource for every Revit user:
http://images.autodesk.com/adsk/files/autodesk_revit_2014_model_performance_technical_note.pdf

Application Notes: 3ds Max Design

Autodesk 3ds Max Design has base system requirements that are about the same as they are for Revit. However, 3ds Max Design stresses your workstation differently and exposes weakness in certain components. With 3ds Max Design there isn't any BIM data interaction to deal with, although linking RVT / FBX adds a lot of overhead. Instead, 3ds Max Design 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 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:

- Polygons - 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 Design is a highly tuned and optimized multi-threaded application across the board. Geometry, viewport, lighting, materials, and rendering subsystems can all be computationally expensive and 3ds Max Design will take full advantage of multiple cores / processors. Having many fast cores allows for fast interaction with the program even with very large scenes. 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. Using CPUs that have 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.

All told, 3ds Max Design can make whole use of the best CPU you can afford. If you spend a lot of time in 3ds Max Design and render high resolution images, you owe it to yourself to look at more highly-powered workstations that feature two physical multi-core CPUs. 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 also requires a lot of system memory, particularly for large complex scenes with Revit links as well as rendering operations. The application itself will consume about 640MB without any scene loaded. If you regularly deal with large animation projects with complex models and lots of textures, you may find the added RAM capability found in very high end workstations - upwards of 192GB - to be compelling in your specification decisions. The choice of CPU decides how much RAM your system can address, due to the internal memory controller. Normal desktop CPUs top out at 32GB, and most scenes can readily work fine within this maximum. However, for those who regularly work with large complex scenes, moving to a 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 are faster with a number of improvements accelerating navigation, selection and viewport texture baking. Apparently anti-aliasing can be enabled with minimal impact on performance but real-world experience says this largely depends on the graphics card.

A big differentiator in graphics platform selection is the rendering engine used. Unlike mental ray, the iRay rendering system can directly use the GPU for rendering tasks to a very high degree. This obliquely plays into the choice of CPU as this determines the number of PCI Express lanes, so if you want 3, 4, or even 5 graphics cards to leverage in iRay, you necessarily need to specify a high-end CPU and a hardware platform that can handle multiple graphics cards. We specifically discuss the needs of iRay users in 3ds Max in the section on graphics hardware.

Storage: The 3ds Max Design 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 3ds Max artists 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 two main features are the Clash Detective (in Navisworks Manage only) that identifies and tracks collisions between building elements before they are built, and the TimeLiner which applies a construction schedule to the building elements, allowing you to simulate the construction process. Navisworks 2015 adds integrated 2D and 3D quantification for performing easy takeoffs.

As such, Navisworks is all about fast viewpoint processing as you interactively navigate very large and complex building models. Most of these have been extended from the Design Intent models from the design team to include more accurate information for construction. These kinds of construction models can be from various sources outside of Revit, such as Fabrication CADmep+ models of 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, 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 2015 we now have the multi-threaded Autodesk 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 Design. But because Navisworks does not need the same kind of application-specific information stored within Revit, performance on very large models 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. However, the new Autodesk rendering engine in Navisworks 2015 will demand higher performance systems to render effectively. If you are planning to do rendering from Navisworks, target your system specifications for Revit and 3ds Max Design.

RAM: Navisworks 2015 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. 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.

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 Design 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 may suffer from some "1.0" newishness. It 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.

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 - sometimes hours depending on how many there are. The indexing operation is heavily reliant on the CPU and disk as it writes out the (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 2015 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 130MB by itself without any drawing files loaded. ACA weighs in at 180 MB, and AMEP at 214MB. 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 is an application that graduated from Autodesk Labs' Project Newport. Originally designed as a review platform for product industrial design, Showcase provides real-time interaction with ray-traced lighting and materials, allowing you to fluidly visualize your design and make comparative, intelligent decisions faster. While it is not meant for photorealistic rendering, walkthrough animations, or lighting analysis - those tasks are best left to 3ds Max Design – it fulfills the need for a fast, realistic, interaction with your design models.

Now bundled in the Building Design Suite, 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. I tend to think of Showcase as your Project PowerPoint.

Processor: Showcase very much likes a fast CPU to import / load files and handle its primary operations. It can be a slow program to use with large models.

RAM: Showcase consumes a mundane 322MB 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 whopping 770MB. Expect even higher high memory usage with your models.

Graphics: As it relies on DirectX 9 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, 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.

However, it has its own quirks, most of which deal with its relatively slow display performance and somewhat iffy stability. Showcase places great stress on the graphics card; running it alongside Revit, Inventor, and AutoCAD has often caused slowdowns in those applications as Showcase sucks all of the life out of the graphics subsystem.

Section II: Hardware Components

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 3GHz 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. However, there are dozens of CPU models out there, some 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.

This section will discuss four primary kinds of Intel CPUs: The latest 4th-generation Haswell line of mainstream desktop CPUs, the Haswell-E “Extreme Edition” lineup, the Haswell EP Xeon E3 / E5 v3 families, and the latest 4th 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, and other factors like chipsets, memory, and expansion capabilities that will factor into your decision making process.

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[1].” 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, and Haswell. Future microarchitectures will be Skylake, 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.

The term “**Manufacturing Process**” or just “**Process**” describes the way in which a CPU is manufactured. Process technology primarily refers to the size of the lithography of the transistors on a CPU, and is discussed in terms of nanometers (nm).

Over the years we’ve gone from a 65nm process in 2006 with the Pentium 4, Pentium M and Celeron lines, to a 45m process with Nehalem in 2008, to a 32nm process with Sandy Bridge in 2010, and to a 22nm process with Haswell in 2012. In 2015 we should see Broadwell and Skylake ship using a 14nm process, then 10nm in 2016, 7nm in 2018 and 5nm in 2020. With each die shrink, a CPU manufacturer gets more chips per silicon wafer, resulting in better yields and lower prices. In turn we get faster processing using much less power and heat.

[1] http://en.wikipedia.org/wiki/Microarchitecture

The Tick-Tock Development Model

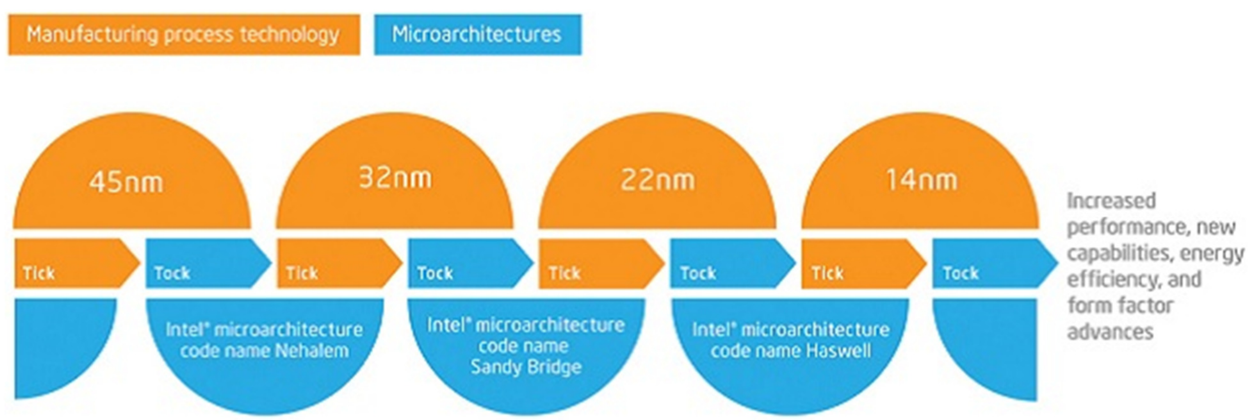
To balance the work between microarchitecture and process advancements, Intel adopted a "Tick-Tock" development strategy in 2007 for all of its future processor development cycles. This strategy has every introduction of a new microarchitecture be followed up with a die shrink of the process technology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

In short, a "Tick" shrinks the process technology used in the current microarchitecture. Shrinking a process is very hard and a big deal, because if it were easy we'd already be at the smallest process possible. Intel pretty much has to invent ways that they can adequately shrink the process and still maintain cohesion and stability in a CPUs operation.

Ticks usually include small but important tweaks to the CPU cores as well, but 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 (which equates to less heat and noise), along with bug fixes, new instructions, internal optimizations, and slightly higher performance at lower prices.

Because these refinements to the microarchitecture may be profound, each die shrink Tick also gets new code names which could be considered a new microarchitectures as well. For example, the Westmere "tick" was not simply a 32nm die shrink of the Nehalem microarchitecture, but added several new features. Ivy Bridge was a 22nm die shrink of 32nm Sandy Bridge, and Broadwell will be a 14nm die shrink of Haswell, if and when it gets here.

Conversely, a Tock is the introduction of an entirely new microarchitecture CPU design based on that smaller process. This is introduced after Intel formally vets the smaller process and has everything working. Every year there is expected one tick or one tock, with some variations in between.



Source: Intel

Legacy CPUs: Nehalem, Sandy Bridge, and Ivy Bridge

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

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 i5 / i7 CPUs of this generation were the first quad-core processors which provided a 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 in a future section.

Nehalem / Core i7 also reintroduced Hyper-Threading, a technology 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.

Basically, Hyper-Threading solves the grocery store checkout line problem. Imagine you are in line at the grocery store and the person in front of you has to write a check, or gets someone to perform a price check. You are experiencing the same kind of blockages CPUs do. Hyper-Threading is what happens 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

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 a BIM / Viz workstation. Core i5 CPUs are quad-core but do not include Hyper-Threading, so they are 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 to Nehalem and represented an impressive increase in 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 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 chances are good that 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 or 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 desktop-oriented lineup included the largely ignored 4-core Core i7-3820 with 10MB of L3 cache, and the 6-core $550 Core i7-3930K and the $1,000 i7-3960X with 12/15MB cache respectively. The introduction of an “extreme” variant will also carry forward with each new microarchitecture.

SB-E was also 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 SB-E is really a die-harvested Sandy Bridge-EP. While the EP-based Xeon will have 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. Sandy Bridge-E CPUs did not include the onboard GPU – 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 that had 8 memory slots.

Another plus for the emerging GPU compute market was the inclusion of 40 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 between PCIe 2.0 x8 and x16.

However, SB-E’s PCIe 3.0 was implemented before the PCIe 3.0 standard was ratified, meaning that they were never fully validated. 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 4 lanes of x4. For people who needed additional GPUs for high end 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, a 16-lane fully validated PCIe 3.0 controller, and relatively small improvements in speed (~ 5-10%), but with a remarkable lowered power draw.

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 compares to the ATI Radeon HD 5850 and NVIDIA GeForce GTX 560, both respectable cards for BIM given Revit's fairly mundane system requirements. For 3ds Max and Showcase, however, 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 die harvesting 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 more 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 - closed loop water cooling is mandatory for IVB-E. 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.

As Haswell and its Extreme variant Haswell-E are the latest and greatest CPUs out there, we will get into the specifics of these chips to a later section.

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 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, current consumption, and temperature, as well as 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, the processor’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, power, 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 standard 4-core CPU, a notation of 8/8/9/10 means that with 3 or 4 cores active, the maximum turbo bin is 8, with 2 cores active the max turbo bin is 9, and with only one core active the maximum 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. For example, the i7-4770 has a base frequency of 3.4GHz and a max Turbo frequency of 3.9GHz. Its Turbo bins are 3/4/5/5. This equates to the following:

No. of cores active	No. of Turbo bin steps	Calculation	Max frequency
4	3	3400 + (3 × 100) = 3400 + 300 = 3700	3.7 GHz
3	4	3400 + (4 × 100) = 3400 + 400 = 3800	3.8 GHz
2	5	3400 + (5 × 100) = 3400 + 500 = 3900	3.9 GHz
1	5	3400 + (5 × 100) = 3400 + 500 = 3900	3.9 GHz

Turbo bins for processors are model dependent and are provided at intel.com and http://en.wikipedia.org/wiki/List_of_Intel_Core_i7_microprocessors. In general, normal desktop processors have low bin numbers because the core base clock is relatively high, leaving lower headroom at the top. Bins of 2/3/4/4 are common in the latest Haswell CPUs. Low powered S and T processors have higher turbo bins (6/9/11/12 for an I7-4770T 2.5GHz) because the base clock is much lower and thus the headroom to up the clock is greater.

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 need boosting. 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.

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. However, 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 get from Dell or HP.

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

Sockets and Chipsets

No talk of any processor is complete without discussing 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 creates a socket standard and a matching PCH to go with it, in various flavors that cater to a certain market.

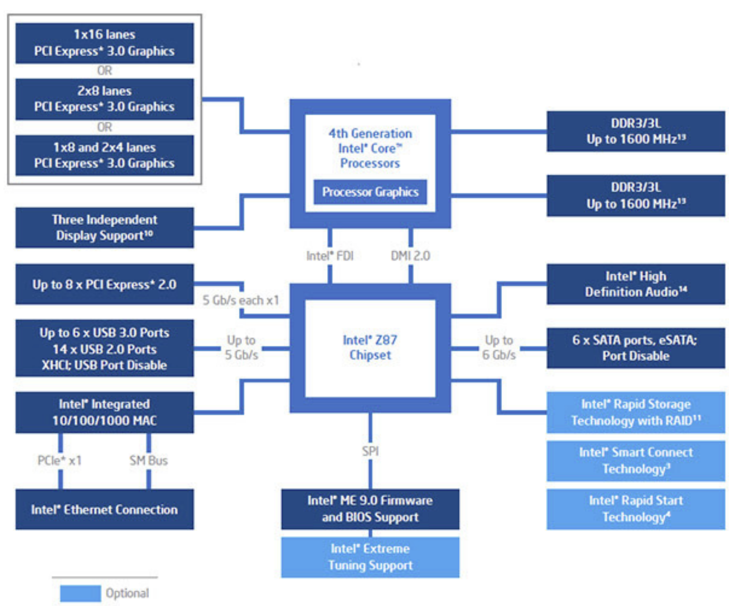
All Haswell CPUs use the **LGA 1150** socket, also called Socket H3. Sockets are typically only important to consumers as far as upgradability goes. LGA 1150 will probably be used by Haswell’s successor Broadwell as well, but not by Skylake. LGA 1150 is the successor to LGA 1155, used in the Sandy Bridge and Ivy Bridge CPUs, meaning you cannot upgrade from one of those to Haswell by simply dropping in a new CPU. However, cooling systems for LGA 1155 are compatible with LGA 1150.

Historically chipsets are the most disappointing aspect of any CPU architecture. Up until recently Intel had a habit of providing brain damaged chipsets which did not meet the needs of the enthusiast in some way. 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.

The original Haswell CPUs (the i7-4770, i7-4770K, and i7-4771) utilized the fully featured 8-series Z87 chipset. (Sandy Bridge used the 6-series Z68 and Ivy Bridge the 7-series Z77). “Z” chipsets are geared toward the enthusiast and have the most features. Intel creates other less powerful chipsets (the H87, H81, and B85 Express) targeted for different markets than ours. We also have “X” series chipsets reserved for the “–E“ processors, e.g. X79 for Sandy Bridge-E & Ivy Bridge-E and the modern X99 chipset used for Haswell-E.

The PCH is really made of one chip, and has high speed interfaces to the CPU, which include Flexible Display Interface (FDI) and Direct Media Interface (DMI) in mainstream non-Xeon models. FDI is included only when the CPU has integrated graphics. In the Xeon E5 family, we have Quick Path Interconnect (QPI), a very fast bus that connects more than one physical CPU together with the PCH.

The PCH’s primary function is to provide second tier I/O to devices such as USB ports, hard disk SATA ports, Ethernet, and HD audio. It is covered by a separate heat sink on the motherboard since it does so much work.



The Z97 Chipset

The Haswell Refresh, or Devil's Canyon CPU uses a new 9-series Z97 chipset. Its primary advantage over the 8-series is to provide more storage options, including SATA Express and the new M.2 standard for internally mounted expansion cards such as Wi-Fi, Bluetooth, and SSDs. M.2 provides 4 PCIe lanes and one SATA 3.0 6 Gbps port, allowing the use of both PCI Express and SATA storage devices in the form of small M.2 cards.

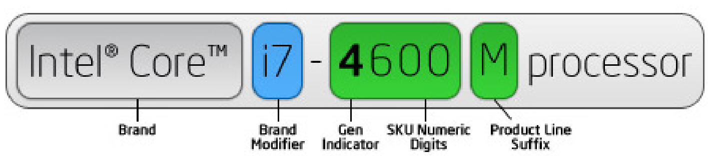
Intel® 9 Series Chipsets

- Supports 4th Gen and 5th Gen Intel processors (Socket LGA1150)
- Intel® RST for PCIe Storage
  - Up to 67% faster than SATA Gen 3 – same speeds as SATA Express
  - Targeted for high-end gaming and data productivity platforms
- Intel Device Protection Technology with Boot Guard†
  - Protects boot block against malware attacks
- Intel® HD Graphics Support
  - Up to three independent displays (w/Haswell CPU)
  - Digital display (HDMI/DVI/DP/eDP) and VGA
- I/O Port Flexibility (SATA, USB* 3.0, PCIe*)
- Intel® Small Business Advantage with 5 MB FW Sku (H97 Only) †
- Intel® Rapid Storage Technology 13†
  - Dynamic Storage Accelerator
  - UEFI Fast Boot Support
  - SATA 6 Gb/s support
  - RAID 0/1/5/10
  - Intel® Rapid Start Technology
  - Intel® Smart Response Technology
  - TRIM on RAID 0 SSD Configurations
- Integrated Gigabit Ethernet MAC

Key Feature Differentiation	Z97	H97
Processor Support	LGA 1150	LGA 1150
Firmware Sku Support	1.5 M	1.5 M 5 M
Graphics Support	1x16 or 2X8 or 1X8 + 2X4	1x16
Independent Displays	3	3
Mem/DIMMs per Channel	2/2	2/2
CPU Overclocking	Yes	No
Intel® Rapid Storage Technology	Yes	Yes
Dynamic Storage Accelerator	Yes	No
Intel® Device Protection Technology with Boot Guard	Yes	Yes
Intel® Smart Response Technology	Yes	Yes
Intel® Small Business Advantage	No	Yes
USB total (Maximum USB3.0)	14(6)	14(6)
IO Port Flexibility	Yes	Yes
Total SATA (Max # of 6Gb/s)	6 (6)	6 (6)
Maximum PCI Express* 2.0 (5GT/s)	8	8
Intel® RST for PCIe Storage	1x2	1x2

Haswell on the Desktop: The Core i7-477x Series

Before we get into the specific models to consider, first let's review how Intel names their desktop CPUs:



Haswell desktop CPUs will use the i7-4xxx name with possibly a processor suffix. Suffixes are:

Product Line Suffix	Description	Example
K	Unlocked CPU multiplier (up to 63x)	Intel® Core™ i7-4770K Intel® Core™ i5-4570K
QM	Quad-Core Mobile	Intel® Core™ i7-4820QM Intel® Core™ i7-4720QM
S	Performance-optimized lifestyle (Low power with 65W TDP)	Intel® Core™ i7-4770S Intel® Core™ i5-4550S
T	Power-optimized lifestyle (Ultra low power with 34-45W TDP)	Intel® Core™ i7-4770T Intel® Core™ i5-4570T

We can remove the QM mobile, S, and T entries from consideration, as they are lower powered variants used for specific purposes. "K" processors with unlocked CPU multipliers allow the chip to be overclocked by tweaking the core clock multiplier in the BIOS.

This leaves us four models to choose from (Source: http://ark.intel.com/compare/75123,77656,80806,80807)

Intel Haswell Lineup				
Processor Name	**Core i7-4770K**	**Core i7-4771**	**Core i7-4790**	**Core i7-4790K**
Code Name	Haswell	Haswell	Haswell	Devil's Canyon
NewEgg.com price:	$334.99	$314.99	$309.00	$299.99
Essentials				
Launch Date	Q2'13	Q3'13	Q2'14	Q2'14
Cache	8 MB			
Bus Type	DMI2			
System Bus	5 GT/s			
Instruction Set Extensions	SSE 4.1/4.2 AVX 2.0			
Price	Box: $350.00 Tray: $339.00	Box: $320.00 Tray: $314.00	Box: $312.00 Tray: $303.00	Box: $350.00 Tray: $339.00
Performance				
# of Cores	4			
# of Threads	8			
Processor Base Frequency	3.5 GHz	3.5 GHz	3.6 GHz	4 GHz
Max Turbo Frequency	3.9 GHz	3.9 GHz	4 GHz	4.4 GHz
Turbo Bins	2/3/4/4			
TDP	84 W			
Memory Specifications				
Max Memory Size	32 GB			
Memory Types	DDR3-1333/1600			
# of Memory Channels	2			
Memory Specifications				
Max Memory Size	32 GB			
Memory Types	DDR3-1333/1600			
# of Memory Channels	2			
Max Memory Bandwidth	25.6 GB/s			
ECC Memory Supported	No			
Graphics Specifications				
Processor Graphics	HD Graphics 4600			
Base Frequency	350 MHz			
Max Dynamic Frequency	1.25 GHz	1.2 GHz	1.2 GHz	1.25 GHz
Video Max Memory	1.7 GB			
Intel® Quick Sync Video	Yes			
Intel® Wireless Display	Yes			
Intel® Flexible Display Interface	Yes			
Intel® Clear Video HD Technology	Yes			
# of Displays Supported	3			
Expansion Options				
PCI Express Revision	3.0			
PCI Express Configurations	1x16, 2x8, 1x8/2x4			
# of PCI Express Lanes	16			

Haswell Performance Improvements

Compared to Ivy Bridge, Haswell cores can execute about 5-10% more instructions in each clock cycle (possibly more when AVX2 instructions are factored in). Haswell also brings its voltage regulation circuitry into the CPU, allowing for faster, fine-grained control over the delivery of power to the CPU cores. This allows Turbo Boost to not only respond more quickly to changes in processor usage, but it allows individual cores to run in different-powered states. This allows Turbo Boost to work at the core level, raising and lowering power to each one independently.

General Notes on Haswell CPUs:

- Clearly, the new Devil's Canyon i7-4790K represents today's sweet spot and best buy in a mainstream desktop CPU. While almost exactly the same in specifications to the other three, it's the newest, fastest and least expensive. It effectively replaces the i7-4770K as the "go to" desktop processor.
- Overclocking: Only the K variants have the processor multiplier unlocked and are available for overclocking. However, testing indicates the overclocking is more limited on Haswell as you have a lower range of voltages to work with to get stable systems. The i7-4770K and the new i7-4790K overclock exactly the same: speeds in the 4.7GHz range seem to be pretty effortless, especially when married to decent motherboards with advanced pushbutton overclocking features. This translates into about a 9-10% decrease in rendering times.

TSX Instructions

The biggest *potential* news with Haswell for Building Design Suite users 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. TSX was to be made available on certain Haswell models, including the i7-4790K.

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. Of course, Autodesk would have to code in TSX support.

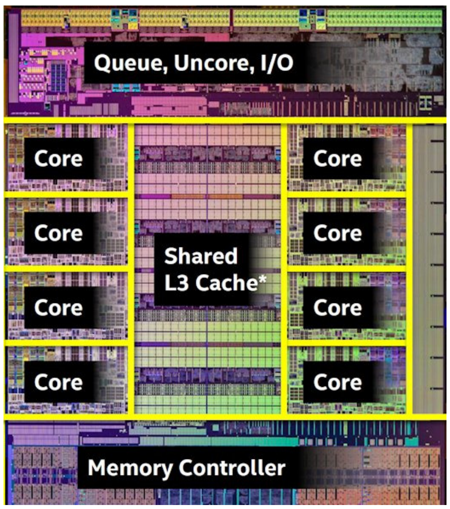
Unfortunately, in August of this year Intel confirmed a serious erratum (bug) found with the TSX instructions that could result in unpredictable system behavior. **Intel actually disabled TSX entirely in all Haswell CPUs, even across workstation and server platforms.** At the time of this writing it is not known when new CPUs would be released with fixed a TSX instruction set.

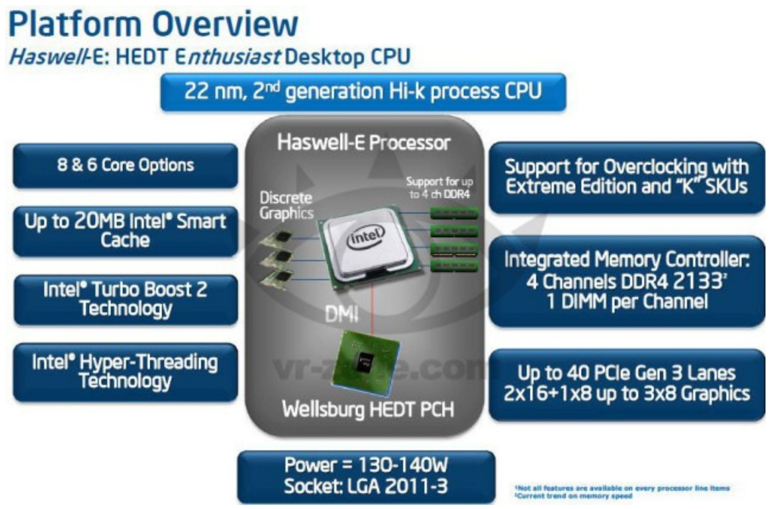
Haswell-E – The Enthusiast's Choice

No generation is complete without a powerful Extreme version, and that is what we got in 2014 with the Haswell-E. Touted as the very best consumer performance CPU available today, Haswell-E is fabricated as the 6-core Core i7-5820K, 6-core i7-5930K, and a true 8-core i7-5960X.

At 2.6 billion transistors, Haswell-E is the first desktop processor to include full support for DDR4 memory, with a quad-channel memory controller, up to 20MB L3 cache, 28-40 lanes of PCIe 3.0 connectivity, Hyper-Threading, and 6 or 8 full cores. This is a very nice uptick from the SB-E and IB-E processors which had 4 to 6 cores depending on model.

In addition, all Haswell-E processors are unlocked, meaning they are open to overclocking by enthusiasts.





Source: Intel

Intel Haswell-E Processor Lineup			
Processor Name	**Core™ i7-5820K**	**Core™ i7-5930K**	**Core™ i7-5960X**
Cache	15 MB	15 MB	20 MB
Pricing	Box : $396.00 / Tray: $389.00	Box : $594.00 / Tray: $583.00	Box : $1059.00 / Tray: $999.00
# of Cores	6	6	8
# of Threads	12	12	16
Processor Base Frequency	3.3 GHz	3.5 GHz	3 GHz
Max Turbo Frequency	3.6 GHz	3.7 GHz	3.5 GHz
TDP	140 W		
Max Memory Size	64 GB		
Memory Types	DDR4-1333/1600/2133		
# of Memory Channels	4		
Max Memory Bandwidth	68 GB/s		
ECC Memory Supported	No		
PCI Express 3.0 Lanes	28	40	40

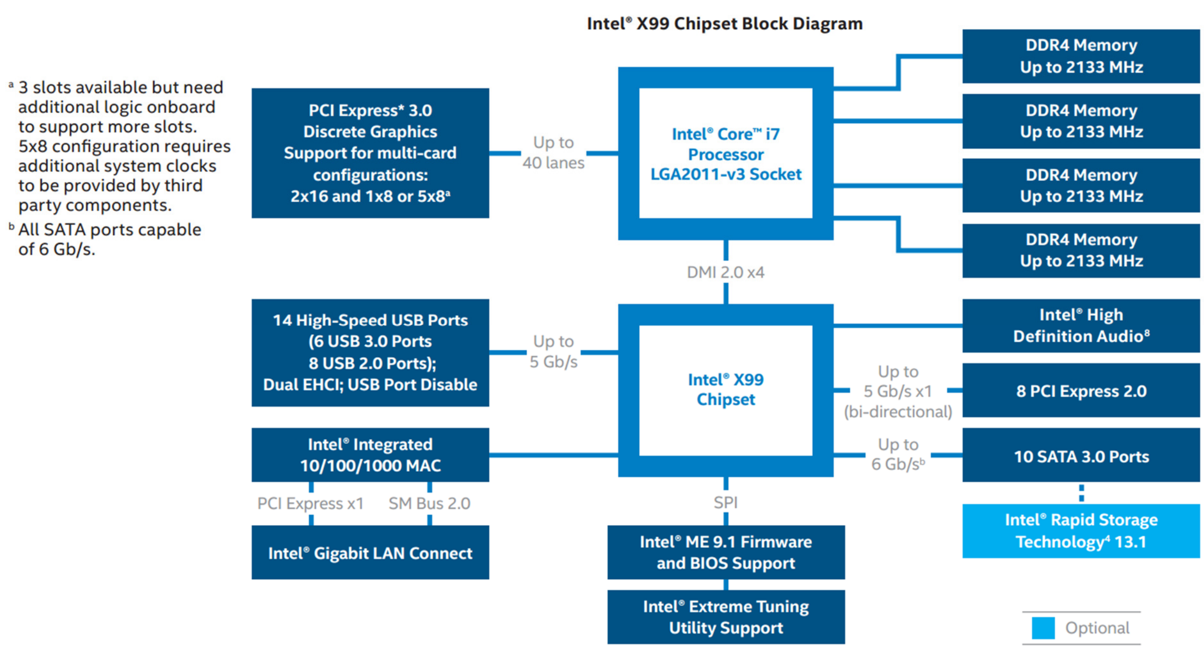
Notice that the top of the line 8-core i7-5960X is only running with a base clock of 3 GHz, which is 25% slower than the Devil's Canyon i7-4790K which runs at a base clock of 4 GHz. This is directly due to the TDP (Thermal Design Power) limitation; to get all 8 cores cooking along any faster would produce too much heat. The highest base clock that would provide lowest TDP was chosen. As with Ivy Bridge-E, water cooling is a must, and Intel offers the same TS13X closed loop solution.

But it's **FAST**. At stock speeds and benchmarking multithreaded applications, the Haswell-E is faster than the fastest Ivy Bridge-E. The Core i7-5960X even stomps on the Devil's Canyon i7-4790K, despite the 1GHz deficiency in clock speed: In most benchmark tests that stress multi-threading, there is roughly a 45 percent to 50 percent improvement going from the Haswell to the Haswell-E part. The differences get tighter when comparing last year's Ivy Bridge-E six-core Core i7-4960X but it's still a nice, big number.

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 it would only be limiting if you planned to stuff more than 3 GPUs into the case.

With 28 lanes, dual GPU setups would run at x16/x8, rather than x16/x16. Three-GPU setups on a 28-lane chip would run at x8/x8/x8 instead of the x16/x16/x8 they would run with 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 compute performance for low money. Note that motherboard manufacturers can leverage all 40 PCIe lanes allowing up to a 5x8 configuration using 3rd party components.

Haswell-E is mounted in the LGA 2011-3 socket and supported by the newer X99 chipset, which fixes the sins of the X79 chipset by supporting plenty of USB 3.0 ports and up to 10 SATA ports @ 6Gb/s. Typical PCH fare such as 8 PCIe 2.0 lanes, Gigabit networking, and audio are included. The only glaring omission is the lack of official SATA Express or M.2 support that came with Z97.



Xeons

The Xeon line covers workstation and server applications, and there is a broad allocation of cores, speeds, and cache to fit almost any need. Intel's Xeon processors traditionally live much longer shelf lives than their desktop cousins. Models which are several generations old are still available in today's high end workstations, so you need to shop carefully to get today's hotness.

Xeons differentiate themselves from desktop CPUs with larger caches, support for ECC memory, exotic 8/10/12/18 core counts, quad-channel memory controllers and support for much greater installed RAM sizes. Some have the ability to run more than one on a motherboard, allowing 2-way and even 4-way configurations.

Xeons are typically found in professional-class machines such as the Dell Precision and HP Z series workstations. While you can build one yourself, it will be expensive. These rarefied components are not sold on the open market with the same volume as typical desktop components, and you as a single buyer do not have the purchasing power of a Dell or HP, so all of the primary components - CPU, motherboard, and graphics card - will likely cost much more than a packaged system.

Haswell-EP

Today's Xeons are based on the Haswell microarchitecture. Higher-end server / workstation derivatives are based on Haswell-EP, which adds a slew of additional features on top of what is in Haswell or Haswell-E. Haswell-EP is not a single chip, but a family of three, segmented by the number of cores it contains. All three chips are fabricated on the latest 22nm process technology and all share the same basic DNA. With Haswell-EP Intel has segmented them by core count, with quite a bit of separation in terms of size and transistor count between the three options.

Code name	Cores / Modules	Threads	Last-level Cache Size	Process Node	Estimated Transistors (millions)	Die Area (mm^2)
Haswell-EP	8	16	20 MB	22 nm	2601	354
Haswell-EP	12	24	30 MB	22 nm	3839	484
Haswell-EP	18	36	45 MB	22 nm	5569	662

Xeons offer the following improvements over their desktop cousins:

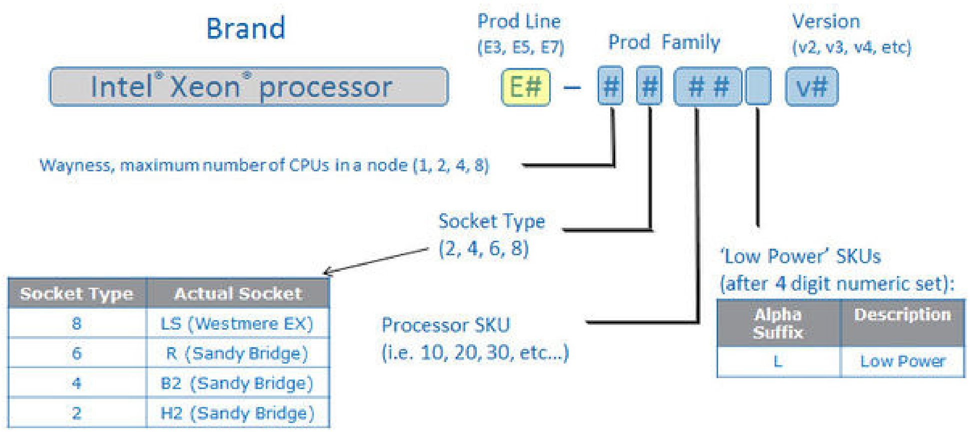
1. Incorporation of voltage regulation (VR) circuitry directly onto the CPU die. Called FIVR for "fully integrated voltage regulator," this allows for more efficient operation, as voltage transitions with FIVR can be much quicker than with an external VR. FIVR has many more supply lines, allowing for fine-grained control of power delivery across the chip. FIVR allows individual cores to maintain its own power state, or P-state, with its own clock speed and supply voltage. This also reduces the physical size of the CPU and its support circuitry.
2. A more efficient Turbo Boost algorithm that controls the chip's clocking behavior. The power control routine now monitors the activity of each core for throughput and stalls. If it decides that raising the clock speed of a core wouldn't be energy efficient—presumably because the core's present activity is gated by external factors or is somehow inefficient—the Turbo mechanism will choose not to raise the speed.
3. Better virtualization performance with decreased time to needed to enter and to exit a virtual machine. Haswell-EP also allows 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.

4. Haswell-EP raises the bandwidth on the QPI socket-to-socket interconnect to 9.6GT/s, up from 8GT/s before. The PCIe 3.0 controllers have been enhanced with more buffers and credits, so they can achieve higher effective transfer rates and better tolerate latency.

The Xeon Product Lines

Xeons are split into three main product lines. The E3 series is the entry-level end of the spectrum, basically Xeon versions of Haswell desktop processors, and are meant for entry-level workstations. The E5 represents the largest segment of Xeons with 29 different models. It powers up the serious graphics workstation as well as mid-class servers. The E7 is its own weird 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 going to buy an E7 based system, Facebook itself probably has more than a few.

The naming convention for Xeons is markedly different from desktops but are highly informative. It follows the nomenclature of E#-#### v#, described in this graphic:



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 it uses. The last two numbers are the specific product SKU, where the numbers ending in zero or one do not have an 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 (v2, v3) is arguably one of the most important codes as it tells you what microarchitecture the processor is based on. The version-less parts are based on Sandy Bridge, V2 parts were based on Ivy Bridge, and V3 parts are based on Haswell.

Xeon E3-12xx v3 Family

For the purposes of this discussion, we are going to limit ourselves to considering only v3 Xeons, that is, those based on the Haswell microarchitecture, as anything else is last year's technology and unworthy of concern. As stated previously, members of the Xeon E3 family are meant for entry level workstations and are really Xeon versions for corresponding desktop CPUs, with very similar specifications. The difference is that while there may only be one worthy desktop model (i.e., the i7-4790K), Intel builds plenty of usable Xeons in this class, separated only by 100 MHz incremental clock speeds.

The specific lineup is the E3-12xx v3. They work in single CPU only configurations and use the LGA 1150 socket. As with their desktop cousins, they come with 4 cores (some with Hyper-Threading, some not, so be careful), a dual-channel memory controller supporting up to 32GB RAM, 8MB cache, support for the new TSX instructions, and so on. Aside from the available clock speeds and support for ECC RAM, they are almost identical to the desktop Core i7.

Intel makes 24 or so different E3-12xx CPUs, so once we filter out those that (a) don't have an IGP, (b) aren't low powered versions, and (c) support Hyper-Threading, we get these 8 contenders:

Processor Name	Launch Date	Cores / Threads	Base Frequency	Max Turbo Frequency	L3 Cache	TDP	Bus Type	Bulk Price
E3-1230 v3	Q2'13	4 / 8	3.3 GHz	3.7 GHz	8 MB	80 W	DMI	Tray: $240.00
E3-1231 v3	Q2'14	4 / 8	3.4 GHz	3.8 GHz	8 MB	80 W	DMI2	Tray: $240.00
E3-1240 v3	Q2'13	4 / 8	3.4 GHz	3.8 GHz	8 MB	80 W	DMI	Tray: $262.00
E3-1241 v3	Q2'14	4 / 8	3.5 GHz	3.9 GHz	8 MB	80 W	DMI2	Tray: $262.00
E3-1270 v3	Q2'13	4 / 8	3.5 GHz	3.9 GHz	8 MB	80 W	DMI	Tray: $328.00
E3-1271 v3	Q2'14	4 / 8	3.6 GHz	4 GHz	8 MB	80 W	DMI2	Tray: $328.00
E3-1280 v3	Q2'13	4 / 8	3.6 GHz	4 GHz	8 MB	82 W	DMI	Tray: $612.00
E3-1281 v3	Q2'14	4 / 8	3.7 GHz	4.1 GHz	8 MB	82 W	DMI2	Tray: $612.00

Source: http://ark.intel.com/compare/75056,80908,75057,80907,75052,80910,75055,80909,75054

Notes:

1. Remember that we had the Haswell refresh called Devil's Canyon in Q2 2014 which upped the base frequency after Intel solved the thermal interface issue. Here we can see the newer Xeons (shaded) introduced in Q2'14 that have adopted the updated Devil's Canyon chip in the E3-12x1 v3 models, which have their speeds bumped up 100MHz in speed and are the same price as their predecessors. *In other words, don't consider any Xeon E3 model ending in a 0.*
2. Pricing is in 1000-bulk trays to system integrators. Box prices are on average about $9-10 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 from the E3-1271v3 @ 3.6 GHz to the E3-1281v3 at 3.7 GHz. Paying $284 more for a tiny 100MHz speed increase is just dumb.

Conclusion:

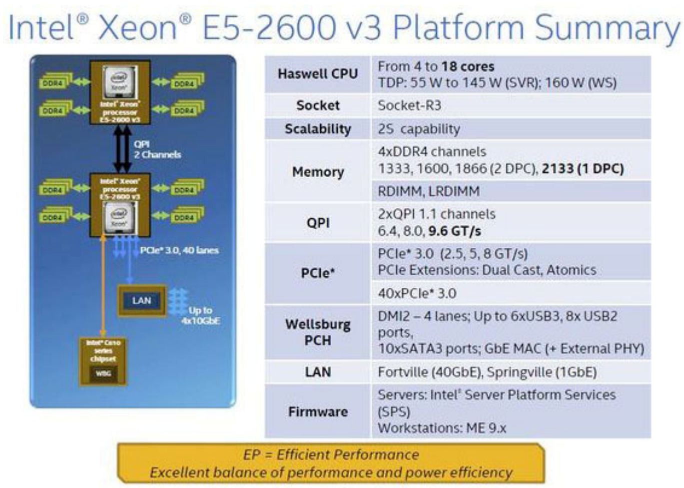
For the E3 family, I think the **E3-1241v3** @ 3.5 GHz at $262 holds the sweet spot right now. However, remember that the flagship desktop Devil's Canyon CPU, the i7-4790K, is running at a 4 GHz base clock for $350. That's only $88 more than the E3-1241 v3 for a huge 500MHz jump in speed. The i7-4790K also handily beats the E3-1271K in benchmark tests. ***If your system vendor offers a Devil's Canyon CPU alternative in your workstation specification, opt for it instead of a Xeon E3-12xx v3.*** The two CPUs are so close as to be considered functionally identical. Since you don't want or need ECC memory, the E3 is not a compelling CPU given the existence of the i7-4790K.

Xeon E5 v3 Family

The Xeon E5 is different. It represents the broadest implementation of Intel's workstation / server class CPU and is easily a highly recommended starting point for any BIM workstation. This family has models specifically tuned for everything from medium- and high-end graphics workstations to high-end database and multiuser Citrix / Cloud servers. From four to 18 cores and up to 3.6GHz base clock rates, and thermal designs range from 55W to 145W on the server side, and as high as 160W for the highest-end workstation part. Overall the sheer number of SKUs is impressive; Intel actually uses three different dies to create all of the different CPU models in the E5 family.

As mentioned previously, the Xeon E5 v3 is built on the Haswell-EP stepping of the Haswell microarchitecture, similar to the Haswell-E discussed previously. However, the –EP variant has the capability to scale up to a bewildering 18 cores and uses a ring technology to feed data to all of the cores. As with other high-end CPUs, Xeon E5 v3s have 40 PCIe 3.0 lanes and 4-channel memory controllers supporting up to 768GB of DDR4 RAM.

The Xeon E5 v3 is split into two main product lines. The E5-16xx v3 is a single-CPU configuration Haswell-EP, and uses the more mainstream Direct Media Interface (DMI) to connect the CPU to the PCH circuity. DMI tops out at about 5GB/s, which is not terribly fast, but enough for the downstream PCIe 3.0, USB, SATA, and other low-speed device interfaces that are hanging off of the PCH.



The E5-26xx v3 can be implemented in dual-CPU systems. It uses a very fast Quick Path Interconnect (QPI) to connect the two CPUs together as well as connect to the chipset. This point to point interconnect was first introduced back in 2008 in the Bloomfield line of high end Core i7-9xx series of desktop CPUs, which were some of the fastest CPUs on the market. QPI provides enough bandwidth - up to a stunning 25.6GB/s - to connect multiple processors together and to provide high speed PCIe devices from the chipset instead of (or in combination with) the on-CPU PCIe controller. This allows more high-bandwidth PCIe 3.0 devices, which is great for multi-GPU rigs or multiuser systems.

Xeon E5-16xx v3: For workstation use we are interested in models run at a base clock speed of at least 3.0 GHz. The four 4 & 6-core models you would look at are listed below. Unlike some Ivy Bridge-E Xeons of yesteryear, all of the v3 Haswell-EP units support Hyper-Threading.

Processor Name	Launch Date	Cores / Threads	Processor Base Frequency	Max Turbo Frequency	L3 Cache	TDP	Bulk Price
E5-1620 v3	Q3'14	4 / 8	3.5 GHz	3.6 GHz	10 MB	140 W	$294
E5-1630 v3	Q3'14	4 / 8	3.7 GHz	3.8 GHz	10 MB	140 W	$372
E5-1650 v3	Q3'14	6 / 12	3.5 GHz	3.8 GHz	15 MB	140 W	$583
E5-1660 v3	Q3'14	8 / 16	3.0 GHz	3.5 GHz	20 MB	140 W	$1080
E5-1680 v3	Q3'14	8 / 16	3.2 GHz	3.8 GHz	20 MB	140 W	$1723

The best deal here is probably the 6-core E5-1650 v3. Running at a base clock of 3.5GHz with 15MB of L3 cache for $583, it's both faster and cheaper than its desktop cousin the Core i7-5930K. On the high end, the E5-1660 v3 is essentially a Xeon version of the high-end 8-core i7-5960X desktop CPU.

Xeon E5-26xx: Capable of being used in dual-CPU configurations, the E5-26xx is a solid workstation and database server processor, with 4, 6, and 8 cores (up to 16, 24, or 32 cores with Hyper-Threading), dual QPI links, and support for 384GB of RAM. However, as the number of cores rises the based clock tends to lowers, so you have to balance the massive multiprocessing capability with the speed required for single-threaded apps.

With over 12 different models to choose from, the five models below are your workstation contenders. Even though clock speed is relatively slow, two 8-core CPUs with their combined 32 Hyper-threaded cores will smack down everyone else when it comes to mental ray rendering.

Processor Name	Launch Date	Cores / Threads	Processor Base Frequency	Max Turbo Frequency	L3 Cache	TDP	Bulk Price
E5-2623 v3	Q3'14	4 / 8	3.0 GHz	3.5 GHz	10 MB	105 W	$444
E5-2637 v3	Q3'14	4 / 8	3.5 GHz	3.7 GHz	15 MB	135 W	$996
E5-2643 v3	Q3'14	6 / 12	3.4 GHz	3.7 GHz	20 MB	135 W	$1,552
E5-2667 v3	Q3'14	8 / 16	3.2 GHz	3.6 GHz	20 MB	135 W	$2,057
E5-2687W v3	Q3'14	10 / 20	3.1 GHz	3.5 GHz	25 MB	160 W	$2,141

Overall the pricing for the 4 and 6 core models is pretty high for what you get – at these speeds, the benefit of the E5-26xx series is the dual-CPU capability. The 8 and 10 core models are crazy expensive, but you would conceivably save the entire cost of at least one other box in a render farm environment.

Intel Mobile CPUs

When talking about outfitting a new laptop with a CPU you have many more things to consider. Laptop models are pretty closed. CPUs consume power and generate heat. Therefore the CPU choice is really subservient to the laptop system as a whole, and you typically choose a laptop platform first, based on size, portability, and power. In any laptop lineup you get a limited choice of CPUs to base it on, because the laptop design itself is largely determining the CPU TDP rating it can safely work with.

Therefore this section is divided into two parts. The Core M represents the low-power, highly efficient required for ultrabooks and other highly mobile platforms. The Core i7 series represents nearly desktop-like performance in a mobile chip, which would be on "mobile workstation" platforms like the Dell Precision M4xxx series.

Core M

Intel's mobile portfolio of CPUs got a new member with their new Core M platform. This is Intel's moniker for the Broadwell-Y series of processors, a follow up from Haswell-Y and the first release of their new 14nm process technology. Recall that we were supposed to get 14nm Broadwell in desktop processors in 2014, but Intel had issues in manufacturing. Instead we got the "do over" with the Devil's Canyon Haswell refresh.

14nm is clearly the next step to deliver solid performance in mainstream x86 applications in the next generation of fanless laptops, tablets, and smartphones. As a 14nm die shrink "tick" of Haswell, the fundamental architecture of Broadwell doesn't change much, except to shrink it to be 36% smaller. The IGP supports DirectX 11.2, OpenGL 4.2 and OpenCL 2.0, at an UltraHD resolution of 3840x2160.

There are seven Core M CPUS as of this writing. They are all dual-core iterations, with the Intel HD 5300 graphics IGP and all running at a tiny TDP of 4.5W. 4.5W is a target benchmark TDP for an 11.6, 8mm thick fanless tablet. While none of these CPUs should be in the running for a mobile workstation meant to be running Revit or 3ds Max Design on their own. Remember what we discussed in the beginning of this handout: Cloud-based computing cycles are available at very low prices, meaning that you may want to consider an ultra-portable platform for lightweight, highly mobile computing that can still connect back to big iron in your data closet or in the cloud for the heavy lifting.

The nomenclature for these processors is a little weird, but essentially it breaks down similarly as the mainstream Core series: The ‘5’ means it is similar to the i5 (i.e., dual-core design with Hyper-Threading); ‘Y’ is for Broadwell-Y; and ‘70’ is a relative SKU number.

As the Core M is making its way into tablets first, the current round of benchmarks are typically judging and entire platform instead of just the chip. Today, Microsoft’s Surface 3 is considered about the fastest tablet out there, about 2x faster than an iPad Air and, being Windows, able to run all x86/x64 applications. The Core M-5Y70 is currently benching slightly faster than the Surface 3 in overall 3DMark 1.2 scores, and about 42% faster in web browsing benchmarks.

	M-5Y70	M-5Y51	M-5Y31	M-5Y10c	M-5Y10a	M-5Y10	M-5Y71
Launch Date	Q3'14	Q4'14	Q4'14	Q4'14	Q3'14	Q3'14	Q4'14
Cores / Threads	2 / 4						
Processor Base Frequency	1.1 GHz	1.1 GHz	900 MHz	800 MHz	800 MHz	800 MHz	1.2 GHz
Max Turbo Frequency	2.6 GHz	2.6 GHz	2.4 GHz	2 GHz	2 GHz	2 GHz	2.9 GHz
L3 Cache	4 MB						
TDP	4.5 W						
Max Memory Size	16 GB						
Memory Types	LPDDR3 1600/1333; DDR3L/DDR3L-RS 1600						
Processor Graphics	Intel® HD Graphics 5300						
Graphics Base Frequency	100 MHz	300 MHz	300 MHz	300 MHz	100 MHz	100 MHz	300 MHz
Graphics Max Frequency	850 MHz	900 MHz	850 MHz	800 MHz	800 MHz	800 MHz	900 MHz

4th Generation Mobile Core i7 Processors

When outfitting a mobile workstation adept enough for daily Revit, Navisworks, and 3ds Max work, the choice of CPU is the paramount design criteria. Get something too slow and you will hate using the laptop. Get something too fast that runs too hot, and you’ll hate using the laptop.

While Haswell is a hit on the desktop, Haswell is really a mobile-focused architecture, and offers mobile versions in the same four primary series as desktops: Core i3, Core i5, Core i7, and Core i7 Extreme processors. The primary differences are that i3 is a dual core CPU without Hyper-Threading, so that’s right out of consideration for our needs. The Core i5 is a quad-core model without Hyper-Threading, so for daily workstation use let’s consider that out as well. That leaves the Core i7 and Core i7 Extreme lineup.

When choosing a mobile platform the processor choice will necessarily need to be a balanced decision between power, weight, affordability, and speed. I would only recommend the i5 if I were considering a thin highly mobile ultrabook platform that needed to run Building Design Suite applications in a supporting role, such as showcasing models to clients, doing light Revit work on the road, or other relatively low impact workouts.

Confusing the issue is the bewildering number of Core i7 mobile CPU SKUs, so it helps to be able to strip out those that don't cut the high fidelity cheese. We are going to narrow down the field by choosing CPUs that are true 4-core designs with Hyper-Threading and run a respectable speed with a base clock in the 2.5-3.0 GHz range.

The current crop of Haswell Core i7 mobile processors are named i7-4xxxMQ or i7-4xxxHQ, with the differences between the MQ and HQ almost inconsequential. The HQ is physically about 75% shorter than the MQ, allowing it to be housed in smaller, lighter laptops. The HQ's HD Graphics 4600 can scale up fractionally faster than the MQ. If you are given a choice between an HQ and MQ of the same clock speed, choose the HQ, but don't be put off if the system only offers MQ chips.

On top of all of this, Intel introduced a new Level 4 cache technology dedicated to a newer Iris Pro graphics 5200 IGP in order to boost memory bandwidth and lower latencies. This Haswell variant with better graphics is called Crystal Well and is expected to provide a solid leap in performance from the integrated graphics chip

This means that if you are considering an ultrabook or smaller laptop that doesn't have a discrete graphics card, you will want to look for models with Crystal Well. If you are going to get a discrete card, it won't make any difference.

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

	i7-4710MQ	i7-4800MQ	i7-4810MQ	i7-4900MQ	i7-4910MQ
Launch Date	Q2'14	Q2'13	Q1'14	Q2'13	Q1'14
Cache	6 MB	6 MB	6 MB	8 MB	8 MB
Bulk Price	$378.00	$378.00	N/A	$568.00	$568.00
Cores / Threads	4/8				
Processor Base Frequency	2.5 GHz	2.7 GHz	2.8 GHz	2.8 GHz	2.9 GHz
Max Turbo Frequency	3.5 GHz	3.7 GHz	3.8 GHz	3.8 GHz	3.9 GHz
TDP	47 W				
Max Memory Size	32 GB				
Memory Types	DDR3L-1333,1600				
Processor Graphics	Intel® HD Graphics 4600				
Graphics Base Frequency	400 MHz				
Graphics Max Dynamic Frequency	1.3 GHz				
# of Displays Supported	3				

And here are the Crystal Well variants along with the Extreme Edition models:

	i7-4870HQ	i7-4960HQ	i7-4980HQ	i7-4930MX	i7-4940MX
Code Name	Crystal Well	Crystal Well	Crystal Well	Haswell Extreme Ed.	Haswell Extreme Ed.
Launch Date	Q3'14	Q4'13	Q3'14	Q2'13	Q1'14
Cache	6 MB	6 MB	6 MB	8 MB	8 MB
Bulk Price	N/A	$623.00	$623.00	$1096.00	N/A
Cores / Threads	4/8	4/8	4/8	4/8	4/8
Processor Base Frequency	2.5 GHz	2.6 GHz	2.8 GHz	3 GHz	3.1 GHz
Max Turbo Frequency	3.7 GHz	3.8 GHz	4 GHz	3.9 GHz	4 GHz
TDP	47 W	47 W	47 W	57 W	57 W
Max Memory	32 GB	32 GB	32 GB	32 GB	32 GB
Memory Types	DDR3L-1333, 1600	DDR3L-1333,1600	DDR3L-1333, 1600	DDR3L-1333,1600	DDR3L-1333,1600
Processor Graphics	Iris™ Pro Graphics 5200	Iris™ Pro Graphics 5200	Iris™ Pro Graphics 5200	Intel® HD Graphics 4600	Intel® HD Graphics 4600
Graphics Base Frequency	200 MHz	200 MHz	200 MHz	400 MHz	400 MHz
Graphics Max Frequency	1.2 GHz	1.3 GHz	1.3 GHz	1.35 GHz	1.35 GHz
# of Displays Supported	3	3	3	3	3

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 push a desktop CPU's gaming prowess over everything else, which emphasizes certain properties in the video subsystem, much differently than the 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.com, understand what it is: It is an average of submitted benchmarks from users of 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 and are skewed.

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

Processor Model	Cores / Threads	CPU Speed (Base - Max Frequency)	Passmark – CPU Mark Score
Lynnfield Core i7-860 (c.2009)	4 / 8	2.8 - 3.5 GHz	5,088
Sandy Bridge Core i7-2600K (c.2010)	4 / 8	3.4 – 3.8 GHz	8,584
Ivy Bridge Core i7-3770K (c.2012)	4 / 8	3.5 - 3.9 GHz	9,628
Haswell “Devil’s Canyon” Core i7-4790K	4 / 8	4.0 - 4.4 GHz	11,291
Haswell-E Core i7-5930K	6 / 12	3.5 - 3.7 GHz	13,497
Haswell-E Core i7-5960X	8 / 16	3.0 - 3.5 GHz	15,984
Xeon E3-1241 v3	4 / 8	3.5 - 3.9 GHz	9,825
Xeon E3-1271 v3	4 / 8	3.6 – 4.0 GHz	10,265
Xeon E5-1650 v3	6 / 12	3.5 - 3.8 GHz	13,299
Xeon E5-1660 v3	8 / 16	3.0 - 3.5 GHz	15,532
Xeon E5-2637 v3	4 / 8	3.4 - 3.7 GHz	10,519
Xeon E5-2687W v3	10 / 20	3.1 - 3.5 GHz	15,848
Xeon E5-2699 v3	18 / 36	2.3 - 3.6 GHz	25,192

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

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 market for this benchmark - the enthusiast gamer - is probably tweaking out their Haswell and Haswell-E DIY systems, which can easily be put together from the open components market.

The Final Words on Processors

The latest Haswell and Haswell-E 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 rule is on the 3ds Max dedicated graphics workstation with dual CPUs.

For most mainstream applications, as well as probably 90% of your Building Design Suite needs, the Haswell Devil’s Canyon CPU will do nicely at an appropriately moderate price point. The **i7-4790K** is fundamentally equivalent to its Xeon cousin the **E3-1271 v3**, but is clocked a little higher. Both provide excellent baseline performance for designers of all kinds.

For high-end Revit and 3ds Max Design use, where your building models are in excess of 300MB and you are constantly working with very large scenes for modeling, animation, and rendering, start with the 6 core Haswell-E processors as a solid workstation base. It’s a particularly good platform and a system built on the middle tier 6-core **i7-5930K** or the high-end 8-core **i7-5960X** can readily be built from parts found online.

For corporate workstations, base it on the Xeon E5 family, specifically the **E5-1660 v3** processor for simple CPU systems or two **Xeon E5-2687W v3s** for high-end 3ds Max use. Outfit it with 32GB or 64GB of memory and a great video card (or three).

Beyond this, if you really need as much horsepower as you can throw at your Revit and Max problems, look at systems with dual physical CPUs on one board, or look at very high core counts. The absolute top of the line is the **Xeon E5-2699 v3**, with 45MB of L3 cache and 18 cores, which rockets away with a Passmark score of 25,192, almost twice as fast as the Haswell-E i7-5930K. While it may perform well in rendering, it probably won't perform well in most anything else, and it rolls along at a pokey 2.3 - 3.6 GHz. Pricing isn't readily available, but $4,000 for the chip would not be unexpected.

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 (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.

System Memory

Memory has always been the least exciting thing to worry about when specifying a new workstation. Back in olden times (2012 or so) you bought 8 or 16GB or so of standard DDR3 DRAM and were happy, dang it. Conventional wisdom would have you specify 8 or 16GB for a basic Revit workstation that doesn't do a lot of rendering, and 32GB for a higher end machine that is expected to take on heavier concurrent loads, such as working in Revit, 3ds Max, and Photoshop all at the same time. But the answer really isn't that simple. There are various memory chip speeds and capacities that factor into the equation.

DDR3 Memory Specifications and Speed

Until we are taken over by Haswell-E and later CPUs, today's workstation memory standard is still a 240-pin DDR3 SDRAM DIMM, which is an abbreviation for Double Data Rate Type Three Synchronous Dynamic Random Access Memory, Dual Inline Memory Module. For laptops and mobile workstations, today's mainstream standard is a 204-pin DDR3 DIMM.

DDR3 DIMMs can be specified and purchased at different speeds, provided as two standard name formats: DDR3-zzzz where zzzz = the data rate in mega- transactions per second, or PC3-xxxx where xxxx = the peak transfer rates in MB/s. DDR3 denotes the data rates of raw DDR3 chips, whereas PC3 denotes the theoretical bandwidth for a complete assembled DIMM.

Each speed has a standard for its internal timings and latencies as defined by JEDEC (Joint Electron Device Engineering Council) which is the governing trade organization standardization body. The specifications you will see in the market today are as follows:

DDR3 Chip Name	DIMM Module Name	Memory Clock (MHz)	Cycle Time (ns)	I/O Bus Clock (MHz)	Data Rate (MT/s)	Peak Transfer Rate (MB/s)	Timings (CL-tRCD-tRP)	CAS Latency (ns)
DDR3-1066E DDR3-1066F DDR3-1066G	PC3-8500	133⅓	7 ½	533⅓	1066⅔	8533⅓	6-6-6 7-7-7 8-8-8	11 ¼ 13 ⅛ 15
DDR3-1333F DDR3-1333G DDR3-1333H DDR3-1333J	PC3-10600	166⅔	6	666⅔	1333⅓	10666⅔	7-7-7 8-8-8 9-9-9 10-10-10	10 ½ 12 13 ½ 15
DDR3-1600G DDR3-1600H DDR3-1600J DDR3-1600K	PC3-12800	200	5	800	1600	12800	8-8-8 9-9-9 10-10-10 11-11-11	10 11 ¼ 12 ½ 13 ¾
DDR3-1866J DDR3-1866K DDR3-1866L DDR3-1866M	PC3-14900	233⅓	4 2/7	933⅓	1866⅔	14933⅓	10-10-10 11-11-11 12-12-12 13-13-13	10 5/7 11 11/14 12 6/7 13 13/14
DDR3-2133K DDR3-2133L DDR3-2133M DDR3-2133N	PC3-17000	266⅔	3 ¾	1066⅔	2133⅓	17066⅔	11-11-11 12-12-12 13-13-13 14-14-14	10 5/16 11 ¼ 12 3/16 13 ⅛

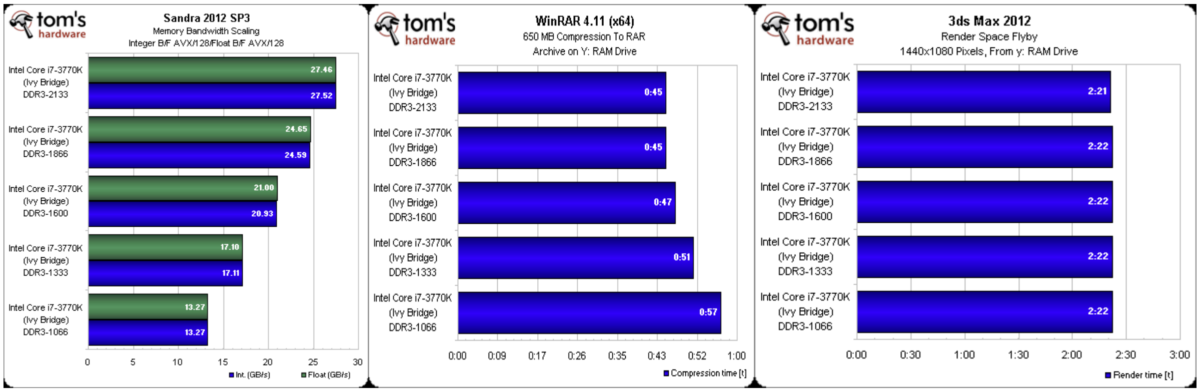
You won't see the DDR3 chip name letter suffix when purchasing RAM; it's simply there to define the standard. The above chart lists the standards specified by JEDEC, so you may see off-the-books memory data rates specified for certain modules that aren't listed above. These are manufacturer optimizations using higher-quality or over-volted chips.

In general you only need to consider the module speed, and even then, don't fret too much. If you are really into overclocking and tuning your memory settings for sub-second speed improvements, you may look at such nerdy things as timings and/or CAS latency numbers.

An exhaustive review of memory timings and their effect on performance is found here: http://www.anandtech.com/show/3851/ and here: http://www.anandtech.com/show/6372/memory-performance-16gb-ddr31333-to-ddr32400-on-ivy-bridge-igp-with-gskill

Most benchmark tests do not find any appreciable difference in enthusiast-class low latency modules over normal ones. Even module speed itself is really not much of an issue. Probably the best articles out there explaining memory speeds and their effect on overall performance are http://www.anandtech.com/show/4503/sandy-bridge-memory-scaling-choosing-the-best-ddr3 and http://www.tomshardware.com/reviews/ivy-bridge-benchmark-core-i7-3770k,3181-10.html. Both studies show that the internal memory controllers found in all modern day processors have dampened the effects of memory speeds on almost all applications.

The benchmarks below bear this out:



Source: http://www.tomshardware.com/reviews/ivy-bridge-benchmark-core-i7-3770k,3181-10.html

Here you can see that, while memory bandwidth (and bandwidth benchmark scores) jumps dramatically from DDR3-1066 to DDR3-2166, applications are generally not memory bandwidth constrained and thus don't really benefit much from increasing speeds far beyond DDR3-1333, which is typically the lowest speed found in today's workstations. Interestingly, rendering nets zero benefit from higher RAM speeds.

With regards to speed, the sweet spot for today's DDR3 seems to be DDR3-1600 (PC3-12800), where the cost increase over DDR3-1333 is small - about $5 - and you do get a minor bump in performance in some applications. The performance gains by going to DDR3-1866 or DDR3-2133 aren't nearly as much.

DDR4 Changes some Rules

The Q2 2014 introduction of 288-pin DDR4 DRAM changes some things, all for the better for Building Design Suite applications. The primary benefit of DDR4 is the higher module density and lower voltage requirements (1.2V vs 1.5 with DDR3), again increasing performance per Watt, coupled with higher data rate transfer speeds to and from the CPU.

DDR4 is expected to move along at 2133MT/s (MT = MegaTransfers) at a minimum, which is the fastest current DDR3 chips can go. In particular, the higher densities possible with DDR4 mean systems will be outfitted with more system memory than is typically done today. Corsair debuted new DDR4 modules rated for 3300MT/s. On the density side, Samsung is exploiting DDR4's native support for die stacking to create high-performance 64GB DIMMs.

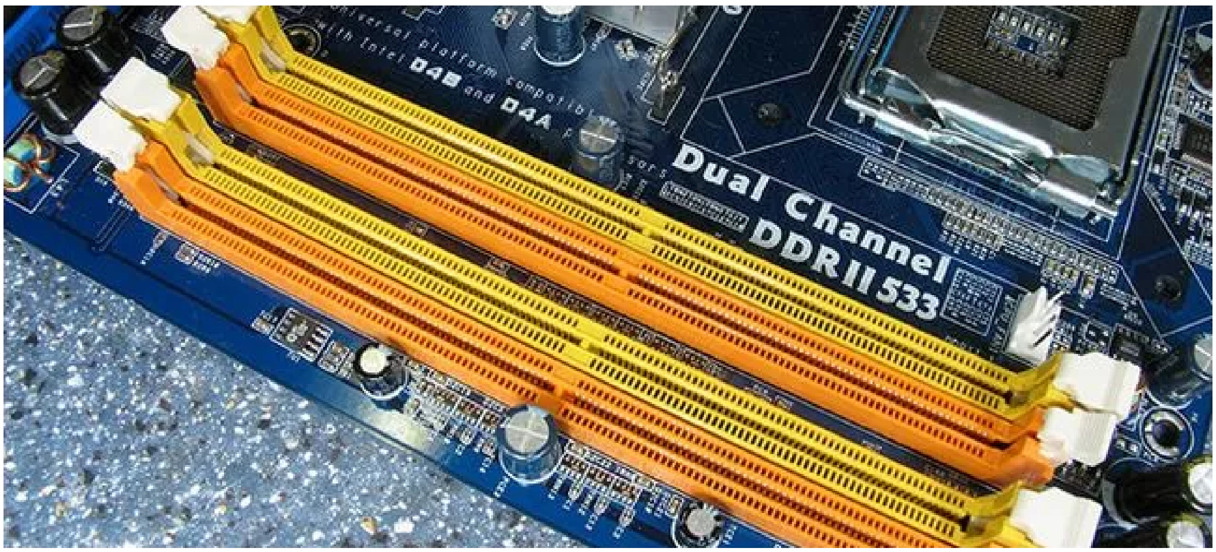
DDR4 is being notoriously slow to gain a foothold in the computing market, because margins are very tight and so much has to change to accommodate it. In short, no one wanted to pay the premium associated with moving to DDR4. With memory controllers now on the CPU, it is the CPU maker (Intel and AMD) that decide how and when this can happen. Then motherboard makers need to step up on their end to implement DDR4 on the motherboard, but that's relatively minor. The slots are keyed differently so you cannot accidentally (or on purpose) install DDR3 in a DDR4 slot.

Current prices for DDR4 are higher than DDR3, but they are clocked much faster. A 4x8GB (32GB total) DDR4 2666 kit is about $500 retail. A similar 4x8GB kit of commonly used DDR3 1600 is about $350 (with a 32GB kit of DDR3 2666 = $450 on up). However, with upper echelon Haswells gaining traction, DDR4 adoption will become more commonplace and prices will appreciably drop.

DIMM Channel Configuration

When configuring memory for your system, it important to understand how a particular CPU works with system RAM. Because different CPUs have different memory controllers, memory configurations aren't all the same. But it's all pretty simple.

The number of channels on a CPU's memory controller tells you how many DIMMs you need to install to get peak performance. DDR RAM works at the faster double-data rate when there are two RAM DIMMs installed per channel, so each channel need to be fully populated for the RAM perform at its highest rate. By doing so the RAM 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 per channel, so you should install RAM in the same-colored slots before moving on to fill others.



Mainstream desktop processors from Sandy Bridge to Haswell, as well as E3-12xx Xeons all contain dual-channel memory controllers, meaning RAM should be installed in pairs and systems typically come with four slots. Older Bloomfield Core i7-9xx processors came with triple-channel controllers and systems had 6 slots. Many system vendors conveniently forgot this fact and shipped triple-channel systems with 2 or 4 DIMM modules instead of the required 3 or 6. The "Extreme" CPUs Sandy Bridge-E / Ivy Bridge-E / Haswell-E as well as all of today's Xeon E5s contain quad-channel controllers where memory is installed in fours, and systems come with 4, 8, 12, or 16 memory slots (8 being most common on workstations, 12+ on servers).

Memory Configuration, Upgradability, and Costs

Regarding configuration cost, recall from the Industry Pressures section on price vs. performance compression where noted that RAM pricing is such that you need to ensure you are optimizing your dollar cost per GB. If you want 16GB on your system you can configure the RAM as 2x4GB DIMMs or 4x4GB DIMMs. RAM prices fluctuate wildly, but today's average 8GB DIMMs are now priced about $75 ($9.375/GB) and 4GB DIMMS are around $50 ($12.5 /GB). In other words, 8GB DIMMS are about 25% cheaper per GB than 4GB DIMMs, so it makes sense to go with the 2x8GB configuration. Bonus: This also gives you 2 slots free on 4-slot systems, so you have the potential to upgrade to 32GB easily. With all four slots filled with 4GB DIMMs, you would need to start over to get to 32GB.

Also recognize that it is always cheaper and better to buy the machine with as much RAM as you can right at the start. Upgrading memory later has the risk that you will have unmatched memory DIMMs which could affect performance and stability. Plus you have the overhead of taking a system down, unplugging everything, installing the new DIMMS, and putting it all back together.

When purchasing RAM for a DIY system, ensure that you get modules with the exact same specifications and batch numbers. Check them carefully when you receive them before they are installed, even if a packaged kit. I once bought a (4) DIMM kit where one of the DIMMs did not have the same batch number of the other three. The system was extremely unstable, rebooting all the time. Troubleshooting the system by swapping out RAM sticks revealed problem down to the one oddball DIMM which had the different batch number. Once replaced, the system has not had an issue since.

ECC Memory

On workstations, you will most likely find options to get Non-ECC or ECC memory. ECC stands for Error Correcting Code, and it is a standard specification that can detect and correct the most common kinds of internal data errors and corruption. Errors can creep in from electromagnetic interference, background radiation, cosmic rays, and other unavoidable things. Errors in memory can lead to system crashes or incorrect data values being generated. ECC memory is specified in systems where data corruption cannot be tolerated under any circumstances, such as the scientific or financial computing circles.

Support for ECC memory is limited strictly to the Xeon lineup of CPUs; it’s 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. System integrators will specify ECC memory in their off-the-shelf preconfigured systems. However, in many cases you can specify non-ECC RAM even on high-end workstations.

For Building Design Suite workstation use, ECC memory may not be as critical. Additionally, current 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.

Graphics Cards

The selection of graphics card for systems using Building Design Suite is problematic, to say the least. There is enough uncertainty about the graphics market to make decisions difficult. It’s actually not that hard, and this section will attempt to dispel some preconceptions you may have about what to look for.

To a large degree, the capabilities and performance of the card is proportional to the cost of the card, but only to a point. What you will see is that some operations such as 2D CAD are almost universally covered adequately across all cards that range from $100 to $5000. It’s how the card handles 3D data that can be an issue, particularly in high end applications like 3ds Max Design and Showcase.

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 or rendering using the iRay renderer. 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 pick a more moderately priced card. 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 the Building Design Suite starts by looking at the two primary kinds of cards being offered. In this corner you have expensive, professional workstation class cards meant for “design professionals,” In the other corner you have a plethora of cheap gaming cards meant for the “great unwashed.”

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 gaming cards. They will call them 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 design used in the Radeon HD 7xxx series desktop boards. The GK104 GPU used in the $1,800 Quadro K5000 is the same one used in the $400 GeForce GTX 780. AMD and NVIDIA fine tune the innards of each for a particular target price point, but the vast majority of the plumbing is the same. Carefully examining the benchmarks proves this out time and time again.

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. Their boards and their drivers are certified by Autodesk, Catia, Solidworks, and a dozen other application developers, as well as system vendors like Dell and HP, to ensure the card meets the minimum specification under each of the applications for which they are tested. This is expensive and costs NVIDIA / ATI money. However, there is clear evidence to suggest that this actually 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. They have special drivers which may take advantage of specific features in those cards meant for high-end content creation, such as antialiasing, z-buffering, and others. Those drivers won't work on the gaming card equivalents.
3. In the past, workstation class cards typically included more onboard RAM, although today this is somewhat ambiguous as gaming enthusiasts demand more video memory. Currently, this is critically important for iRay rendering on NVIDIA-based graphics cards.
4. They have tighter tolerances for build quality. Professional cards are built to run under more rigorous conditions, 24x7x365 inside of server racks, in some cases. To that end ATI and NVIDIA oversee their fabrication with only a small number of manufacturers. This small number of vendors reduces competition which is another reason for higher prices. Gaming cards aren't built to run full bore all day, and the demands of GPU based rendering can really take its toll on the silicon. Such cards have been known to go up in smoke - literally - when rendering many scenes.

While it's a given that professional cards can and do well in Autodesk applications, the strategy by Autodesk, workstation vendors, and others to promulgate professional-class cards to Suite users is hindered by two facts: First, the vast majority of applications in the Building Design Suite are mostly graphics card agnostic. All Autodesk applications use the traditional "gaming" DirectX 9 API as a standard, and any decently armed gaming card has 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 nonsense.

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. But whether you use NVIDIA or ATI cards, you can expect your typical desktop demands to perform roughly the same and have all graphics functions work without tearing, ghosting, or some other weird anomaly.

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

For example: A late model NVIDIA GeForce GTX 970 card costs around $340 and by all accounts will work extremely well for all Building Design Suite application work, including 3ds Max Design. Based on the new second generation Maxwell 2 architecture GM204 chip with 1664 CUDA cores, it's clocked at 1050MHz and has 4GB of RAM. It has a dual-link DVI-I, HDMI 2.0, and 3 DisplayPort 1.2 ports, so it will plug into anything under the sun. By all accounts it represents a solid high-end value. Not the absolute best for everything, but close enough.

On the other hand, an entry-level workstation class NVIDIA Quadro K2200 is based on the 1st generation Maxwell architecture GPU GM107, and comes with 4GB of RAM, 1 DVI, no HDMI, and two DisplayPort 1.2 ports, but only has 1344 CUDA cores. It currently runs around $440 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 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 their design applications. NVIDIA purchased mental ray from mental images back in 2007, and provides Autodesk with the mental ray engine code as they release their versions.

Autodesk takes that finished engine code and simply decides how to implement the renderer in their design applications like Revit, 3ds Max / 3ds Max Design, and Maya. Autodesk then writes the user interface and decides which features in mental ray will be exposed to the user. This is why, although Revit and Max use the same mental ray engine, the interface and capabilities are quite different. Revit's implementation encourages push-button rendering using several presets. Max provides a much more complete under the hood view of all mr settings.

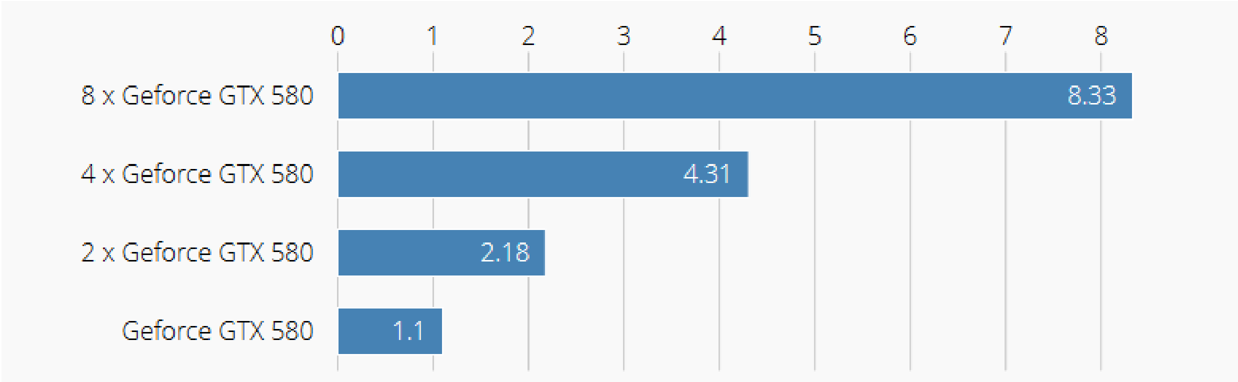
The current version of mental ray used in 3ds Max Design 2015 is 3.12 (in SP2, 3.12.1.17). As additional 3ds Max service packs are released, however, Autodesk will include newer releases of mental ray.

Conversely, iRay was designed by NVIDIA from the ground up to do two major things. First, make it as pushbutton simple to create stunning imagery. Instead of fiddling with literally hundreds of little knobs and controls all throughout the program, the idea is to model up the scene, install the lighting, set the materials, and push the Render button. That's it. Lighting and materials are designed to work together to produce physically accurate results. 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 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. Traditional renderers like mental ray don't use the graphics card for much of anything except to spit the scene out to the monitor, and thus are bound to the speed and capabilities of the CPU.

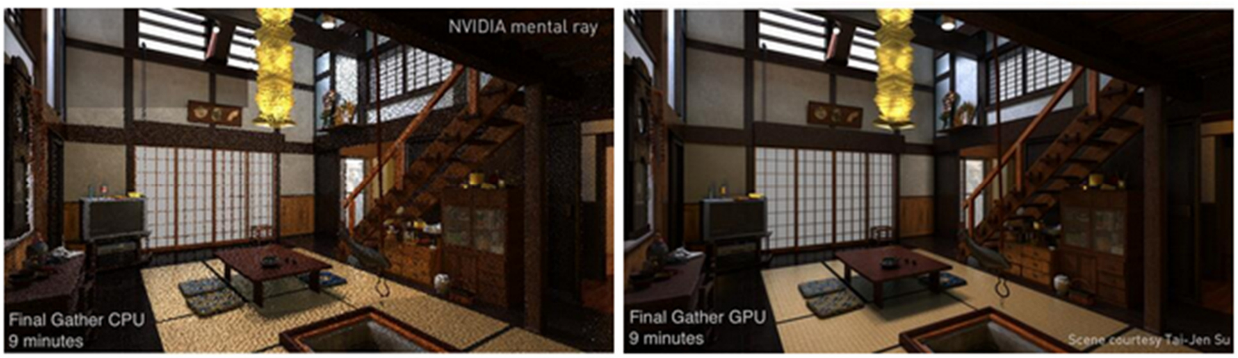
Additionally, iRay rendering performance scales directly with the GPU's CUDA capabilities. Put simply, the number of CUDA cores x the base clock frequency = relative performance. What's more, iRay can leverage and automatically use all of the CUDA cores in your system, even across separate video cards. Add more CUDA GPUs to a system and watch your render times drop. This is one reason why people try to stuff more GPUs into a single system.

Below is a dated chart showing how multiple GPUs in a single system scale up pretty linearly. 2 GPUs is 2x faster; 4 GPUs is 4x faster, and so on.



Source: http://blog.boxxtech.com/2014/11/17/geforce-gtx-rendering-benchmarks-and-comparisons/

NVIDIA has laid out a pretty aggressive roadmap dedicated to GPU accelerated computing in general and GPU accelerated rendering specifically. Even mental ray will be affected, as NVIDIA has stated that all new releases of mental ray will include at least one GPU compute operation. Starting with mental ray 3.11 in 3ds Max 2014, ambient occlusion (AO) can be computed on the GPU – called AO GPU. In 2015 it's still not implemented in the UI just yet, as NVIDIA is still working things out in the lab, but it can be enabled with string options. NVIDIA is also working on GPU acceleration for Final Gather for upcoming releases:



https://twitter.com/mentalray/status/390446138815303680/photo/1

http://blog.mentalray.com/2014/11/27/mental-ray-in-the-lab/

Optix

Along with iRay and mental ray, which are licensed finished products, NVIDIA developed a general purpose do-it-yourself ray tracing engine called Optix. Optix is middleware for ray tracing developers to write their own CUDA custom rendering solutions that just work on NVIDIA hardware. There are several such CUDA-based renderers available out there, such as the highly popular VRay RT, Arion, Octane Render, and others. Other software developers are using Optix and iRay, such as Adobe and Catia, all of which further marginalizing non-CUDA hardware solutions.

2014 – NVIDIA's New Maxwell Architecture

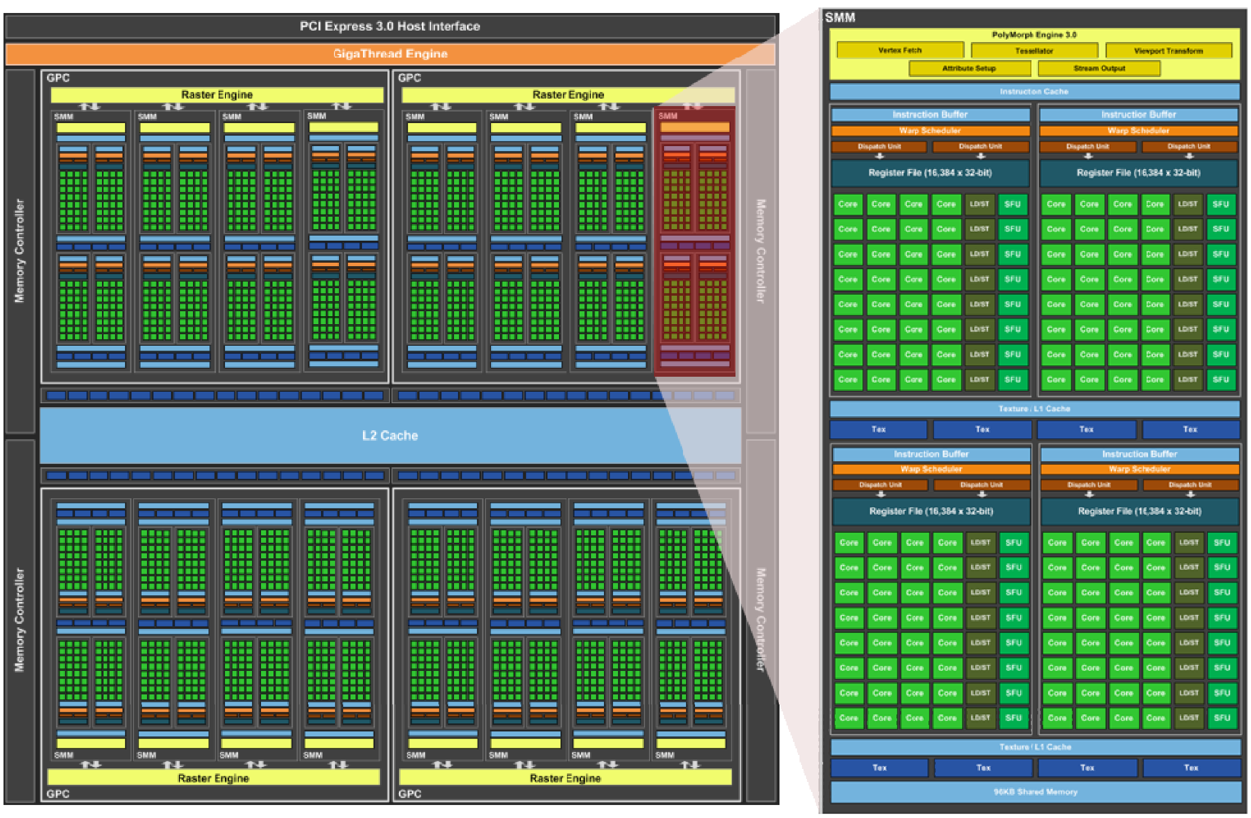
In 2014 NVIDIA debuted the Maxwell architecture, the successor to the highly successful Kepler architecture that powered the GTX 7xx series as well as the GeForce Titan. Maxwell continues NVIDIA's drive to increase Performance per Watt, using the same 28nm process used in Kepler but moving to a smaller node in 2015. In other words, Maxwell is essentially a "Tock" in Intel's Tick-Tock nomenclature.

Maxwell introduced an all-new design for the Streaming Multiprocessor (SMM) that focuses on increasing power efficiency. In terms of GPU architecture, a particular card's performance all revolves around the layout and implementation of the Streaming Multiprocessor, or SMM. Maxwell's new 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 GM207 and GM 208 GPUs. This second generation Maxwell architecture is designed to drive efficiency and raw performance even more, and resulted in the GeForce GTX 970 and GTX 980, which we will look at in more detail later.

Looking under the hood at a block diagram of the newer Maxwell 2 GPU as seen in the GTX 980, 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 SMMs (16 total) and a dedicated raster engine.

Each SMM has 128 CUDA cores, a PolyMorph Engine, and eight texture units. With 16 SMMs, the GeForce GTX 980 ships with a total of 2048 CUDA cores and 128 texture units.



The GM204 Block Diagram and SMM

NVIDIA has also been focused on memory efficiency, both for performance and power reasons, and has increased the L2 from 512KB on GK104 to the same 2MB on GM204. This cache size increase reduces the amount of traffic that needs to cross the memory bus, reducing both the power spent on the memory bus and improving overall performance.

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 fairly slowly compared to other cards as they are often housed in 4U rackmount cases where temperatures need to be low. 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.

Desktop vs. Workstation Benchmarks, Cost Comparisons and Other Factors

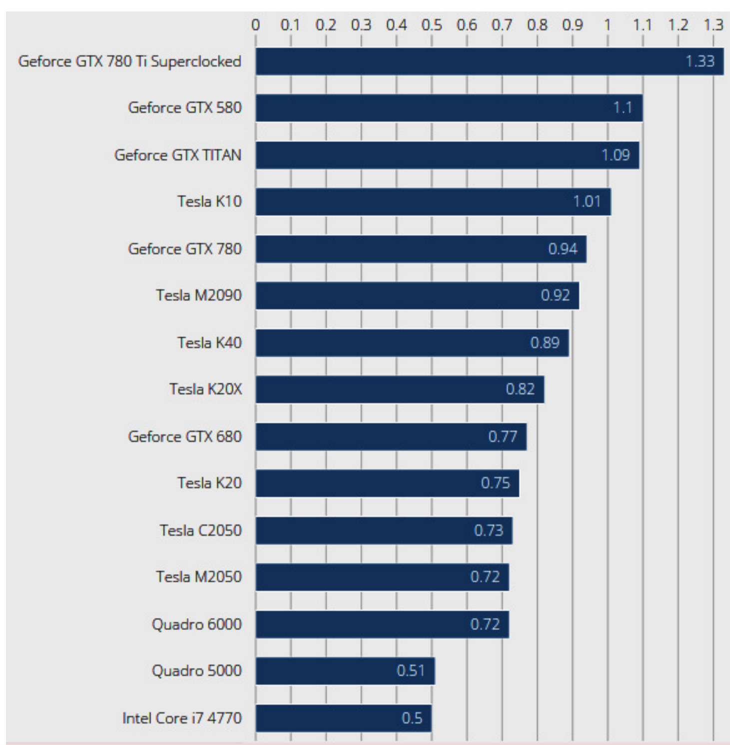
Recall the $100 difference between the less expensive but very capable GTX 970 and the more expensive but mundane K2200. Now look at the super high end, explicitly for CUDA / iRay / VRay RT rendering: A top of the line Quadro K6000 retails for an eye popping $5,000. It runs the "Big Kepler" based GK110 chip with all 15 SMX units and 2,880 CUDA cores operational. Being that CUDA performance scales linearly, we can look at the $/CUDA core as a metric. At $5K the K6000 has a metric of $1.74/core.

Then look at the GeForce GTX 780Ti, one of last year's NVIDIA's flagship gaming cards. It has the exact same GK110 chip with all 2,880 CUDA cores enabled, and is clocked actually higher than the Quadro. It comes in first place in CUDA rendering test and will most likely be at least be as fast as the Quadro K6000. However, it's only $470 – that only $0.16/core. That is over a 10x price difference between cards with almost exactly the same raw capabilities.

In other words, you could conceivably install a GTX 780 Ti, run it until it blows up, get another one, run it into the ground until that one explodes, and repeat the process 10 more times before you've spent the equivalent of one K6000 card. Or, more likely, you could install 3 or 4 GTX 780 Tis in a single system and get compounded performance for the money.

Check out this comprehensive iRay benchmark from www.migenuius.com →

The GTX 780 Ti with it combination of high clock speed and 2,880 cores handily beats every other GTX, Tesla, and Quadro in dedicated iRay rendering benchmarks.



NVIDIA GPU Comparison Chart

Below is a comparison chart which enumerates all of the later model Kepler and Maxwell 2nd generation cards which would be worthy of consideration for Building Design Suite users. It is organized into the 5 current GPUs: 2013's Kepler (GK104), Big Kepler (GK110), 2014's Maxwell 2 (GM204), a GK104-powered Quadro and two GK110-powered Quadros.

NVIDIA GPU Comparison Chart - Part I					
	GTX 770	**GTX 780 Ti**	**GTX Titan**	**GTX Titan Black**	**GTX Titan Z**
Launch Date	30-May-13	7-Nov-13	21-Feb-13	18-Feb-14	25-Mar-14
Architecture	Kepler	"Big" Kepler			
GPU Code Name	GK104-425-A2	GK110-425-B1	GK110-400-A1	GK110-430-B1	2 × GK110-430-B1
SMM / SMX Units	0 / 8	0 / 15	0 / 14	0 / 15	0 / 2×15
Core Configuration [1]	1536 : 128 : 32	2880 : 240 : 48	2688 : 224 : 48	2880 : 240 : 48	2 × 2880 : 240 : 48
Core Clock (MHz)	1046	875	837	889	705
Avg. Boost Clock	1085	928	876	980	876
Pixel Rate (GPixels/s)	33.5	52.5	46.8	53.3	42.3
Texture Rate (GT/s)	134	210	187	213	169
Floating Point (GFlops)	3,213	5,040	4,494	5,121	4,061
Onboard RAM	2GB	3GB	6GB	6GB	2 x 6GB
Memory Clock	7GHz GDDR5	7GHz GDDR5	6GHz GDDR5	7GHz GDDR5	7GHz GDDR5
Memory Bus Width	256-bit	384-bit	384-bit	384-bit	2 x 384-bit
Memory Bandwidth (GB/s)	224	336	288	336	336
Double / Single FP Precision	1/24	1/24	1/3	1/3	1/3
TDP	230W	250W	250W	250W	375W
Transistor Count	3.5B	7.1B	7.1B	7.1B	2 x 7.1B
Manufacturing Process	TSMC 28nm	TSMC 28nm	TSMC 28nm	TSMC 28nm	TSMC 28nm
DX / OGL / OCL / Shader Model	11.0/4.5/1.1/5.0	11.0/4.5/1.1/5.0	11.0/4.5/1.1/5.0	11.0/4.5/1.1/5.0	11.0/4.5/1.1/5.0
Power Connectors	1x6-pin /1x8-pin	1x6-pin/1x8-pin	1x6-pin/1x8-pin	1x6-pin/1x8-pin	2x8-pin
Outputs	2x DVI 1x HDMI 1x DisplayPort	2x DVI 1x HDMI 1x DisplayPort	2x DVI 1x HDMI 1x DisplayPort	2x DVI 1x HDMI 1x DisplayPort	2x DVI 1x HDMI 1x DisplayPort
Newegg Price (Dec. 2014)	$ 314.99	$ 439.99	$ 649.99	$ 1,076.96	$ 1,499.99

[1] Core Configuration = Shader Processors : Texture mapping units : Render output units

NVIDIA GPU Comparison Chart - Part II					
	GTX 970	**GTX 980**	**Quadro K4200**	**Quadro K5200**	**Quadro K6000**
Launch Date	18-Sep-14	18-Sep-14	22-Jul-14	22-Jul-14	23-Jul-13
Architecture	Maxwell 2		Kepler	Big Kepler	
GPU Code Name	GM204-200-A1	GM204-400-A1	GK104	GK110	GK110
SMM / SMX Units	13 / 0	16 / 0	0 / 7	0 / 12	0 / 15
Core Configuration [1]	1664 : 104 : 64	2048 : 128: 64	1344 : 112 : 32	2304:192:32	2880:240:48
Core Clock (MHz)	1050	1127	771	667	902
Avg. Boost Clock	1178	1216	N/A	N/A	N/A
Pixel Rate (GPixels/s)	67.2	72.1	21.6	32	54.1
Texture Rate (GT/s)	109	144	86.4	128	216
Floating Point (GFlops)	3,494	4,616	2,072	3,074	5,196
Onboard RAM	4GB	4GB	4GB	8GB	12GB
Memory Clock	7GHz GDDR5	7GHz GDDR5	5.4GHz GDDR5	6GHz GDDR5	6GHz GDDR5
Memory Bus Width	256-bit	256-bit	256-bit	256-bit	384-bit
Memory Bandwidth (GB/s)	224	224	173	192	288
Double / Single FP Precision	1/32	1/32	1/32	1/24	N/A
TDP	148W	165W	108W	150W	225W
Transistor Count	5.2B	5.2B	3.54B	7.1B	7.1B
Manufacturing Process	TSMC 28nm	TSMC 28nm	TSMC 28nm	TSMC 28nm	TSMC 28nm
DX / OGL / OCL / Shader Model	12/4.5/1.2/5.0	12/4.5/1.2/5.0	11.0/4.5/1.1/5.0	11.0/4.5/1.1/5.0	11.0/4.5/1.1/5.0
Power Connectors	2x 6-pin	2x 6-pin	1x 6-pin	1x 6-pin	2x 6-pin
Outputs	1x DVI 1x HDMI 3x DisplayPort	1x DVI 1x HDMI 3x DisplayPort	1x DVI 0x HDMI 2x DisplayPort	2x DVI 0x HDMI 2x DisplayPort	2x DVI 0x HDMI 2x DisplayPort
Newegg Price (Dec. 2014)	$329.99	$549.99	$818.84	$1,899.00	$4,999.00

Notes on NVIDIA GPUs

1. The GTX 770, based on the “normal” Kepler GK104 GPU, is the oldest of the GPU generations represented here, and has only 8 SMXs and 1536 CUDA cores. While they are great for gaming they may be hobbled too much for extensive use in iRay. However, at $329 you could install one and see how it goes. If you need more CUDA cores, simply install some more in the same system. Note that it completely outclasses the > 2x more expensive Quadro K4200.
2. The “Big Kepler,” based on the GK110 GPU represents the fullest implementation of the Kepler architecture. With 15 SMX units it contains 2880 cores in most models. However, look at the specifics of each card, as NVIDIA will turn off SMX units (and thus remove CUDA cores). With all 15 units enabled it will provide the best iRay performance in a single card.
3. The GTX 780 Ti was last year’s darling and continues to be fantastic for iRay, as it implements all 15 SMXs / 2880 CUDA cores. Its weakness is the 3GB of installed RAM, which could trip up the rendering of complex scenes that won’t fit into memory. Installing multiple GTX 780 Tis in your system will increase your CUDA count and thus iRay speed, but does not mitigate the low amount of

RAM on board. Your results will be directly affected by the complexity of your scene. If your scenes are memory-bound, you would need something like the GTX Titan Z with 6GB or the Quadro K6000 with 12GB.

4. The GTX Titan is last year's model (note the GPU code name) and does not fully implement all of the 15 SMXs on the GK110 chip. It has since been replaced by the Titan Black, this year's version, which has all 15 SMXs (all 2880 CUDA cores) enabled. It is provided for comparison only; do not buy one.
5. The Titan Z is simply two GK110 GPUs on the same physical card. It is essentially two Titan Xs on one card. Because of the increased power load it puts out a lot of heat, and therefore is clocked slower than the Titan Z. It comes with 2x of 6GB RAM which is not shared between the two.
6. The Maxwell 2 GM204 GPU based cards, the GTX 980 and GTX 970, with their newer, more efficient SMM units, perform very well even though they have fewer cores. Increased Performance per Watt means their TDP is a fraction of the Keplers. However, the number of CUDA cores is still paramount for iRay. In benchmarks they perform well but are ultimately outclassed by the more capable GTX 780 Ti and Titan Black. See http://blog.boxxtech.com/2014/11/17/geforce-gtx-rendering-benchmarks-and-comparisons/
7. The Quadro K600 arguably is the best performer in floating point operations (GFlops). However it is very closely followed by the GTX 780 Ti and GTX Titan Black.

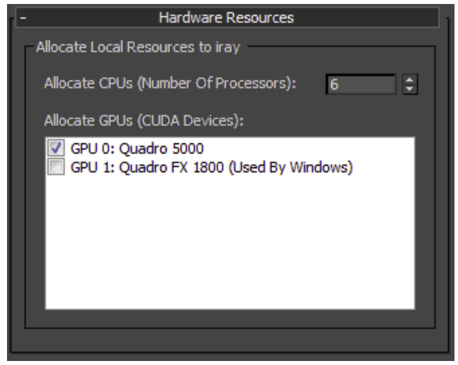
In short, the only reasons to go with a "professional" card are: (a) long term reliability, and (b) the amount of onboard system RAM. This is the one limitation on gaming cards: Currently the 780 Ti has only 3GB of RAM. Currently, when GPU rendering you have to be able to put the entire scene, including geometry and materials, into the card's RAM. If you have multiple CUDA GPUs, each one needs the same amount of RAM because the scene cannot be shared across multiple cards.

However, **this limitation is changing on two fronts.** First, the higher end NVIDIA cards routinely pack 4GB of RAM or more and some, like the Titan Black come with 6GB. The Titan Z, which is essentially two GK110 GPUs on one PCB, comes with 12GB but is packaged as 6GB dedicated to each GPU, and thus are not shared between each other. Second, NVIDIA is actively working on technology that will alleviate this bottleneck. After Maxwell comes Pascal, which will feature stacked DRAM, Unified Memory, and other improvements that should alleviate the memory limitation in iRay.

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. There are limitations to think about when considering GPU rendering.

First, you may have heard of existing technologies for running multiple GPUs, such as SLI (NVIDIA) and Crossfire (ATI). These are technologies meant for gamers who can tie to cards together and aggregate the 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 to handle the Windows OS operations. This doesn't have to be a very high end card - in fact, any decent CUDA 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. Because all systems have a limited number of slots, you can invest in external PCIe enclosures which can house 1 to 4 or more PCIe cards. Remember that 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, if you are using a Mac, a Thunderbolt connection. The main issue with Thunderbolt at this time is driver compatibility and low market usage.

The third reason to put more GPUs in a system is to support more than two monitors. Once you try three monitors you really don't want to go back to only two.

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. The bad news is that no one in the mainstream develops for OpenCL. With VRay RT and iRay being such popular GPU renderers, NVIDIA's CUDA computing platform so fully armed and operational, 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.

Which is a shame, because 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 benchmarks; see http://www.tomshardware.com/reviews/radeon-r9-290-review-benchmark,3659-14.html

Additionally, ATI fundamentally treats GPU compute operations differently from NVIDIA which puts it way out front in other GPGPU operations, such as Bitcoin mining. See https://en.bitcoin.it/wiki/Why_a_GPU_mines_faster_than_a_CPU

PCI Express

So, when considering a system that you know you are going to install multiple GPUs, it pays to understand how PCIe works in your system. When reviewing motherboard specifications you will often see the PCIe expansion slots listed as follows: →

Expansion Slots

PCI Express 3.0 x16	2(x16 or dual x8)
PCI Express 2.0 x16	1(x4 mode)
PCI Express x1	4

This refers to the number of 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.

PCI Express, or **PCIe** for short, is a high-speed serial expansion bus standard that replaced the older PCI, PCI-X, and AGP standards. PCIe makes a whole host of improvements over the older PCI standard, including more efficient signaling, higher bandwidth and throughput, lower physical footprint, and better performance scaling. 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, networking, or whatever.

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 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 slots can each support anywhere from one to 32 lanes in powers of 2 (1, 2, 4, 8, 16, or 32) increments. Lanes are expressed with an 'x' prefix, so x16 represents a 16-lane card or slot which is the largest commonly used size.

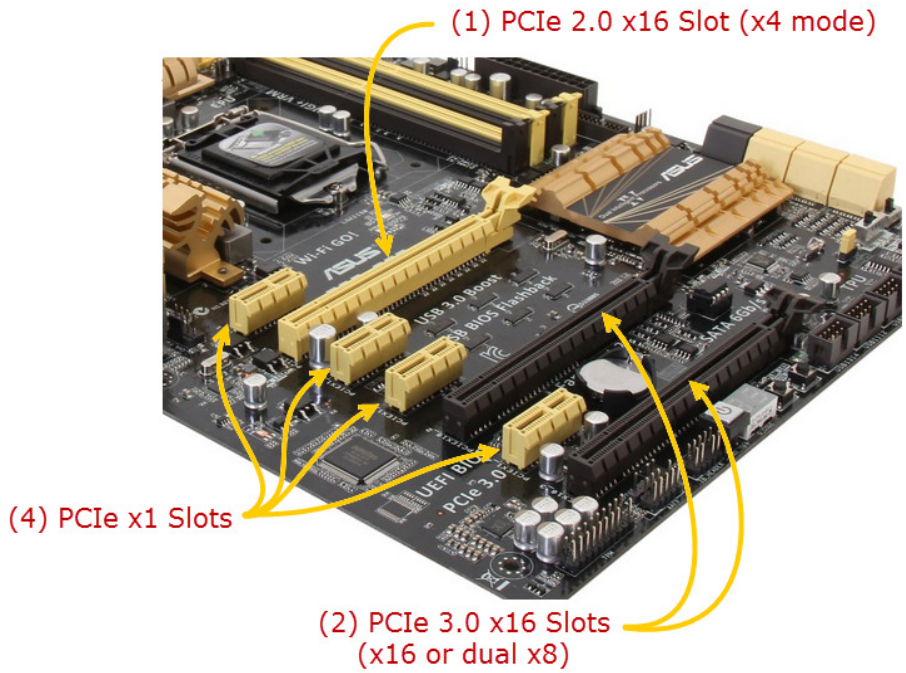
Given this, slots come in different sizes (given in standard x1, x4, x8, x16 notation) and represents the largest physical card that will fit. Smaller cards can fit in larger form factor slots, e.g., you can put an x8-size card in an x16-size slot without issue. However, size alone does not necessarily refer to the slot's bandwidth, sometimes specified as "x**size** (@x**capacity**)," for example, x16(@x8) means an x16 size slot that is configured to only run at x8 speed. Pay attention to your manual motherboard to understand which slot can do what. On most systems, the PCIe slots are colored to indicate their capacity.

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 an 8GB/s aggregate rate per x32 connector. PCIe 2.0 bumped this up to a 500MBs / lane rate (16GB/s aggregate for an x32 connector) and is still commonly used today. PCIe 3.0 doubled this to 985MB/s per lane and is the latest standard. It was first supported by the Ivy Bridge CPU introduced in 2012 and is found in all graphics cards today.

Thus, the slot notation x16, x8, and x4 refers to the number of parallel PCI lanes available to each slot, and based on their PCIe versions represent the amount of bandwidth available to each device. Because the total number of lanes is a physical constant (limited by the motherboard and CPU architecture), the configuration of PCIe slots is always a compromise between throughput and number of devices at that speed.

The Sandy Bridge CPU from 2011 provided 16 PCIe 2.0 lanes. Ivy Bridge and Haswell CPUs provide 16 PCIe 3.0 lanes, and Haswell-E / EP provide 28 to 40 lanes. The various chipsets / PCH that support these CPUs include an additional PCI Express switch which provides an additional set of 8 PCIe 2.0 lanes. The CPU's PCIe 3.0 16 lanes can be configured as (1) PCIe 3.0 ×16, (2) PCIe 3.0 ×8, or (1) PCIe 3.0 ×8 + (2) PCIe 3.0 ×4 slots. The secondary PCIe 2.0 slots can be in x2, x4, and x8 configurations depending on vendor.

The above chart and the motherboard image shown below is for an ASUS Z87-Pro LGA 1150 motherboard, where there are 7 total PCIe slots that divvy up the 16 + 8 lanes as follows: (2) PCIe 3.0 x16 slots which can run in a 1 x16 or 2 x8 mode configuration; (1) PCIe 2.0 x16 slot that runs at x4 mode, and (4) PCIe 2.0 x1 slots. As you can see below, all of the PCIe 3.0 slots are colored dark gray and the PCIe 2.0 slots are yellow.



It is important to understand that the throughput provided by PCIe 2.0 was underutilized by even the most powerful graphics cards, so 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. 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) reduced overall performance by only 14%.

In the preceding section on Processors you can review the differences in the number of PCIe 3.0 lanes available in the various available CPU microarchitectures. With 40 available PCIe 3.0 lanes provided by Ivy Bridge-E, Haswell-E, and Broadwell, users of CUDA-based renderers have more options to use 3-way or 4-way PCIe 3.0 cards at higher capabilities, e.g. 4 cards in a x16/x8/x8/x8 configuration, lowering render times and increasing viewport interactivity.

In addition, some motherboard manufacturers include an additional PLX PEX 8747 chip on desktop systems that increases the PCIe 3.0 lane count by 16, providing 32 PCIe 3.0 lanes overall. Again, this is a solution aimed squarely at users who want 3-way or 4-way GPU setups - high-end gamers or, in our case, iRay users who want to stuff their systems with as many CUDA cores as humanly possible.

Storage

Solid State Drives

Clearly today the biggest news in storage is with Solid State Drives (SSDs). 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 probably 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 - 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 takes a while to load on mechanical drives, but are load very quick from an SSD. The same goes for loading up large Revit models; the speed at which as 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 now under $1/GB for mainstream models - they are nowhere near as cost effective as mechanical drives, where a typical 1TB HD is running around $90, about 9 cents/GB.

Second, how you connect it to your PC can make a big difference. SSDs make full use of the SATA III 6Gbps interface which yields read speeds upwards of 50% faster than on the older SATA II 3.0Gbps interface. Make sure you connect it to the appropriate port on the motherboard.

Enabling AHCI on Legacy Systems

If you upgrade a mechanical drive to an SSD on an existing Windows 7 system, you will probably have to go into the BIOS to change the SATA mode from IDE to AHCI. This will enable Native Command Queuing (NCQ) and TRIM, two SSD-specific features which will improve drive performance and reliability dramatically. You may need to adjust a setting the Registry if you get "`INACCESSIBLE_BOOT_DEVICE`" errors on startup. Refer to this page for the fix: http://support.microsoft.com/kb/922976

Trends and Technical Considerations

In terms of capacity, the go-to sweet spot was typically 240-256GB range, which retails between $170 and $200, depending on drive. This year we see the 480GB segment become much more affordable, ranging around $210 - $250. While 256GB will get you an adequate OS and applications partition, it will be a little tight if you store a lot of additional data on the drive, like music, pictures, and so on. In addition, the internal controllers on this year's models are substantially better, the memory used is faster, and the drives should perform much better.

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 wear down with read and writes 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. To put things in perspective: 1.5PB equates to 50GB of writes / day, for over 80 years.

Today, benchmarks for SSDs show them to be much closer than in previous years. While there may be controller quirks here and there they have largely been solved. However, while all SSDs are orders of magnitude faster than mechanical hard drives, do your research. Some just have better reputations than others. Within a size range, prices are too close to not buy the best.

Benchmarks also show that larger SSD perform better than smaller ones. The performance hit is worse on drives under 240GB. As this is today's most popular size, this should not be much of an issue.

SSD Utilities

Most drives come with a utility, such as Samsung's Magician, that will review the drive's health and optimize your OS for SSD use, such as enable / disable services, write-cache buffers, tweak virtual memory, and set power options. Using your bundled utilities is an important part of making sure the SSD and your system are working to peak performance.

Configuration Options

Because affordable SSDs come in sub-TB capacities, you will probably want to supplement the smaller SSD with additional traditional storage. The best way to incorporate a smaller SSDs is to use it as your C:\ boot drive for the OS and applications, with a larger mechanical D:\ drive for personal near-line storage. A 256GB drive will work fine for most people, as it will easily hold the OS, the complete Building Design Suite Ultimate, an Adobe suite or two, Office 2013, but it will start to get tight. With the 480GB drives coming down to the magic $200 mark, too-small system drives should be a temporary.

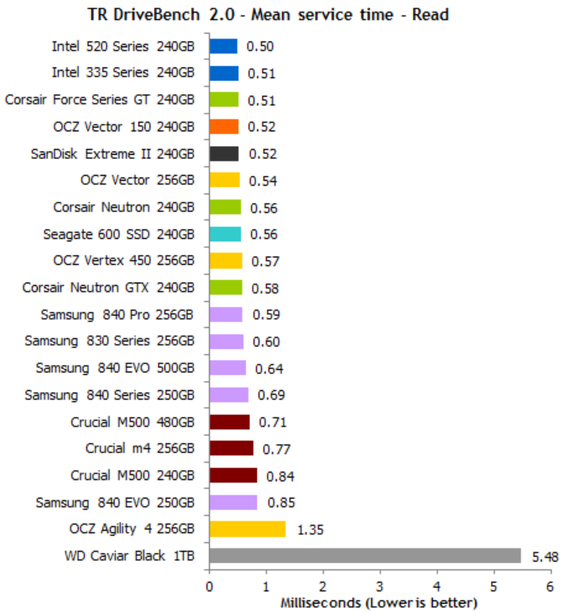
Having a dedicated data drive has the advantage of allowing you to reformat the C:\ drive or reinstall the operating system without hurting the data drive at all. You can also transport the data drive to another system without issue if you house it in a removable enclosure.

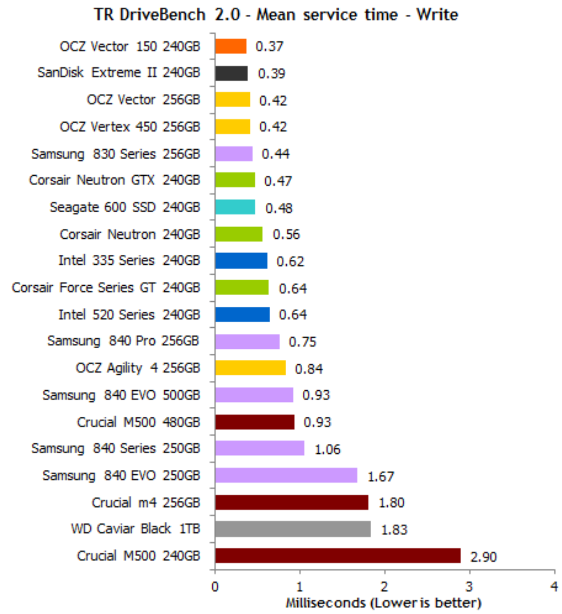
If you are using a very small SSD (e.g., 120GB) and are constantly hitting storage space ceilings, you can save SSD space by putting your virtual memory page file on the D:\ drive. Performance will suffer slightly; page file operations are largely small random reads and larger sequential writes, operations that SSDs crave. However since Windows can multi-thread I/O it should not suffer that much and it can free up a few extra GB. Remember that SSDs don't perform as well when they start to fill up.

If you go this route, set the page file to use the mechanical drive **before** you put any data on it; that will put the page file on the outermost cylinders (the fastest sections of a disk). Set the page file to use the same minimum and maximum space to keep it from fragmenting. However, I would always opt to clean up the SSD first or, if space is constantly a problem, install a larger drive.

Benchmarks

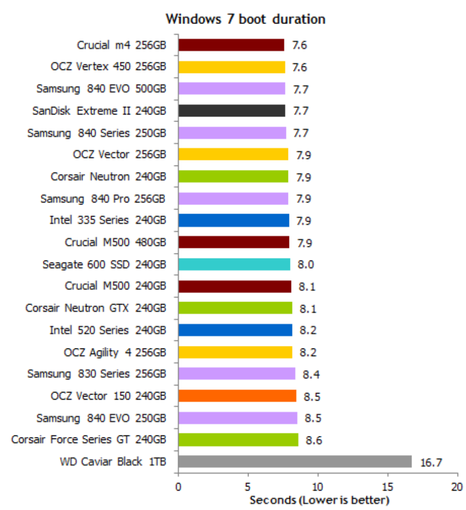
SSD benchmark results show how similar SSD perform against each other in various kinds of operations. As you can see, most modern models do not vary greatly in read speed with others and the overall dynamic range of all SSDs is rather small. Note how every SSD crushes the WD Caviar Black in read speeds, considered one of the fastest mechanical hard drives around. *Write speeds are another matter,* with some performing very well and other not even matching that of a mechanical hard drive, so it pays to research the specific model you want.





Source: techreport.com/review/25610/ocz-vector-150-solid-state-drive-reviewed/2

Given their intrinsically better read speeds, one of the biggest benefits to SSDs is how quickly applications launch and systems boot up:



Source: techreport.com/review/25610/ocz-vector-150-solid-state-drive-reviewed/2

RAID Arrays

One of the ways people try to improve performance is with a RAID (Redundant Array of Inexpensive Disks) system, comprising of two or more hard disks with the data striped across them. This theoretically improves read and writing performance because the system multi-threads I/O to happen asynchronously (at the same time) across multiple devices. However, SSDs have replaced RAID arrays as their performance is much greater.

Today, RAID's main purpose is for data reliability and redundancy. RAID 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 often without turning the server off, meaning downtime is minimized over the life of the server. File, database, and other mission critical servers will most likely use some form of RAID, with 3 or more drives for data security and minimal downtime.

Business workstations 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. Because RAID performance benefits are now eliminated by SSDs, **the only reason to use RAID with SSDs is to aggregate storage**, to get a 500GB volume out of two 250GB drives. 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 wholly unnecessary to purchase special hardware.

Laptops and 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

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 can run Revit 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 every day, 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 with a docking station. 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.

You want to pick a screen resolution for the size of the screen wisely. 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. 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.

Screens can vary wildly in their color fidelity. If you are a graphics professional and need the best color gamut, research the qualities of the particular screens you have to choose from, such as maximum brightness, calibration requirements, contrast, and so on before purchase.

With Windows 8 being so popular the choice of touch screens is now here. Of course, if you are like me, anyone touching my screen is subject to the death penalty, so this may not be a worthy 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. Laptops billed as Mobile Workstations often have 4 slots (sometimes with 2 slots under the keyboard and out of easy reach). Even for intermittent Revit work, I recommend (2) 8GB modules for a total of 16GB to start. As stated before, there’s really no reason to go cheap on RAM when it’s already relatively cheap. A 16GB kit of 2x8GB DDR3-1600 non-ECC 204-pin SO-DIMMs runs about $150.

The same warnings of purchasing RAM upgrades from system vendors applies. 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, 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

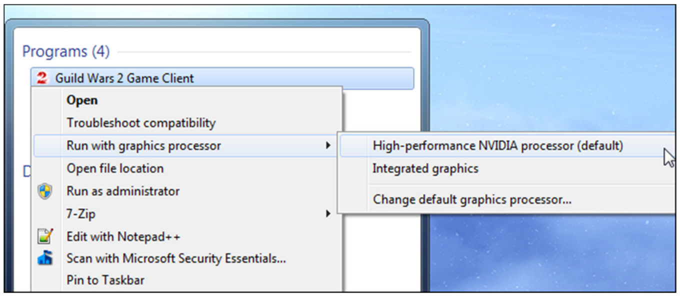
Your choice of graphics card is going to be severely limited within any particular line, so it may end up driving your form factor and model decision. What goes for desktop cards goes for laptops; stay away from the integrated Intel HD graphics unless you rarely touch Autodesk applications.

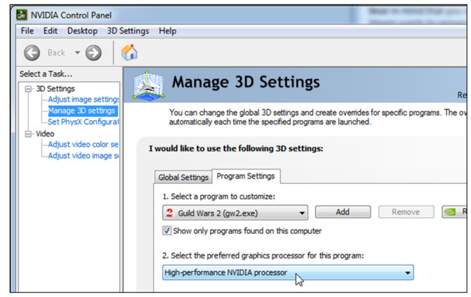
Workstation class laptops are going to come with workstation class graphics, and there’s no option to substitute it out after the fact. In any particular model, I typically recommend upgrading to the fastest option available within reason, because you never know what you may work on next year that really takes advantage of it. 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 GPUs available for a 15.6” platform are the K1100M and K2100M. On the 17.3” platform you have the higher-order GPUs in the K3100M, K4100M, and K5100M. A Quadro K1100M may only be $120 less than a Quadro K2100M, but the K2100M has 50% more CUDA cores (576 vs 384) which could get you through a rendering hump. Of course, it also consumes 10W 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 a K1100 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 one in combination with another drive that is housed in the optical drive bay. A good choice, since most people don't need a DVD drive anymore. You may also have the option for a 1.8" "Mini-Card" that can typically be a SSD, so your choices 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.

If you are storage limited and live with the machine all the time, you will probably hit the 256GB C:\ ceiling pretty early on, so I recommend you opt for a 500GB 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. If you do, you can 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 that you need 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, eSATA, audio, and USB ports. Double check the video ports 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 situation is to have two large 24"+ screens connected to the docking station. By having dual large screens you can keep the laptop closed in the dock. 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.

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. USB 3.0 has largely supplanted eSATA for fast removable storage, as it is just as fast or faster, so you want as many USB 3.0 ports as possible. But you want them in a good location. If you aren't using a docking station you are probably going to plug in a mouse, power cord, and network cable. 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 where my hands naturally rest. Annoying!

You also want a built in high-density SD card reader for transferring files from your camera.

Modern laptops can drive up to three screens, so HDMI is a very important port particularly if you connect to HDTVs or high quality projectors. A direct digital connection to the TV or projector is important to ensure what you see on your high-resolution screen matches the monitor. Older VGA analog connections can drop picture quality and resolution severely.

One of the items on the Dell Precision laptop I use is the numeric keypad. This shifts the keyboard to the left a little and it took some time to get used to it, but it's a necessity now. There's also a dedicated Calculator button which gets a lot of use.

Stuff You Want, You Just Don't Know It Yet

A backlit keyboard is a fantastic feature. It's also a cheap upgrade. 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 the smaller 6-cell one which, with today's power-sipping processors, should last enough to get some things done. Best bet: get a small one for everyday use and a larger backup battery for when you are on the road.

With this, understand your power options and tune them for your typical day. If you stay plugged in all of the time, you can afford to set the power options cranked up for maximum performance. When on a plane, drop it down to maximum battery and turn off things you won't use, like the Wi-Fi transmitter, network jack, and Bluetooth. Dim down the display enough and you can have a battery last an entire flight.

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 Performance Mouse MX 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.

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 losers.

We used to have a fleet of HP laptops for our training lab. From almost every aspect, they were annoying to use. The screen wasn't a high enough resolution. They were dog slow. The keys were in the wrong places and just felt lousy. The cabling for network and power were positioned on the side, and sat right under your left hand, so everything just got in the way.

Then in 2012 we purchased some Dell XPS laptops for our training lab. In general, they were nice machines – they were fast enough and the screen is quite nice. But the keyboard has smaller keys which are easy to miss, making typing a real chore. There are 4 flat buttons in a plastic touch strip along the top, positioned right next to the Delete key, 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 gets 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.

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.

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

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.

The problems of UHD / 4K resolution are pretty severe, enough so that it is very difficult to get a 4K display that actually functions well. The main problem is that graphics cards simply aren’t good enough to spit out that kind of resolution. The second problem is Windows itself. 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 simply isn’t here just yet. 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 Ribbon and other 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. Invest in a DVI-D->HDMI adapter if you do not have an HDMI Out port on your machine. Understand, however, 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 this 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. Look for one with a small USB receiver that can be attached to your laptop or desktop unobtrusively. Longer ones that stick out are prone to getting bumped and can destroy a USB port (ask me know I know).

Personal Recommendations:

Mouse: Logitech Performance MX. I’m on my ~~fourth~~ fifth one. What’s great about this particular mouse is not so much the feel and smoothness, although both are great. It’s 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. 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 (particularly on laptops which often get used 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.

Other types of switches add tactile feedback and varying levels of click noise. Cherry MX Brown switches are is a popular switch with a tactile but non-clicky action. It provides excellent feedback but 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). In addition, Cherry makes a whole host of other color coded switches which are variations of the above, offering stiffer feels, tactile bumps, or quieter operation.

Section III: Buying Guides

Now that we know what components are out there, it is time to start assembling our build. In this section we'll discuss the detailed aspects of purchasing a new system as well as making some targeted suggestions in our 2014 Buying Guide.

Operating Systems

The first major decision to make is the Operating System. Luckily this is pretty easy, as your choices are narrowed down to just a few. In my opinion it's time to forget Windows 7 64-bit and move to Windows 8.1.

Windows 8

2013 saw the widespread adoption of Windows 8 / 8.1 and 2014 cemented it in place. The benefits to the user include:

- Solid performance all around, improved multitasking, and support for more devices and newer technologies like SSDs and USB 3.0
- Extremely quick bootup times due to the hibernating kernel on shutdown
- Excellent and stable Autodesk Building Design Suite 2014 software compatibility
- Fully integrated easy search functionality
- Better desktop user interface, particularly in Windows Explorer
- An overhauled Task Manager which allows much more information and access to system settings
- Extremely quick installation and re-installation of the OS to a clean system
- Integration of Microsoft's SkyDrive cloud storage solution into the OS and applications that support it, particularly Office 2013

Although the biggest "feature" of Windows 8.1, the Metro interface, was universally panned and hated by most, the Windows 8.1 update includes the much needed option to boot straight to your desktop instead of the Metro UI based Start page. It also brought back the Start Button (kind of) but not the Start Menu, which is what people were so angry about in the first place. For those that wish, there are several Windows 7-style Start menu add-ins available like Classic Shell to bring back the older UI completely.

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 systems that can be tested and certified by Autodesk and other software vendors, by specifying 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 in about an hour justifies the warranty.

The Cons:

The primary issues with packaged workstations are:

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

System vendors love creating packaged system designs, 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, the 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 specify 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.

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.

Finally, the ability for you to swap out items is limited. 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. But the fact is that neither Dell nor HP will swap out the Quadro K2000 for a GeForce GTX 980 because it can't get certified. Because vendors can get any parts they want, they *may* allow you to purchase it through them independently, but you would have to install it.

Tips and Tricks for Purchasing Packaged Systems

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

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. **However, always call and talk to a real person. Establish a professional corporate relationship with your vendor and 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 and their warranties and service histories.

Good corporate customers often get freebies such as free shipping or deals on specific components. Dell once shipped me two very large color laser printers as part of a company promotion for their corporate accounts - for free. Of course, they didn't tell us was that the color toner cartridges costs about $400.

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 representative. 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.

I**nclude the physical media for all of the software you are purchasing** when specifying the system. At a minimum this would be Windows 7/8, 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, so having the resource DVDs on hand was a godsend.

Evaluate the extended warranties that will be tacked on to the machine. 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.

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. Some will include additional warranties you did not agree on - you may need to explicitly opt out of them. Out of a hundred quotes that I have received over the years, easily 40% of them had some sort of mistake, such as extra shipping costs, extraneous warranties, and so on.

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. Even if you have a corporate account with Dell or HP, buying parts and assembling machines built for a specific purpose, such as a render farm, can work out well.

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 have some nice rebates 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. Included CPU air coolers will work within the processors temperature envelope, but you may be able to lower operating temperatures by replacing the stock thermal grease with something of quality. 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, and purchasing the stuff you really want separately, such as peripherals or video cards.

Obviously the vendor’s workstation machine will come with a motherboard, case, CPU, a drive, and at least some memory, and probably a chintzy graphics card. Get the fastest CPU you can afford with the least RAM, smallest hard disk, and the lowest end video solution, since those are the parts you want to be specific about. Then purchase the components you really want elsewhere. You may be paying a little extra for parts you don’t need; consider them spares in case anything goes bad and not be completely down.

Tips for 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.

If you buy a complete system from a vendor, **immediately image the drive after you unbox the machine**. Although Windows 7 and 8 make system rebuilds much easier, I will often unpack the machine and immediately yank the drive - before I ever turn the system on - and connect it to another machine using a portable USB 3.0 drive caddy. I’ll use a disk imaging application that will block-level copy the disk structure to a file for safekeeping. If I ever want to get the machine back to an out of the box (OOTB) state, just restore the disk image and its back to pristine shape - great for resale purposes. Vendors like Dell may ship the system with an invisible partition which stores the OOTB image, where you have to enter a special key combination during bootup to access and restore it. Or they install proprietary backup software or some other low-rent solution.

Once imaged and reinstalled, you turn the machine on and accept all license agreements, and get to the desktop for the very first time. **Immediately benchmark the machine** in CPU, memory, and graphics using standard free benchmarks such as SiSoft Sandra 2015, 3D Mark11, PCMark8, Cinebench, and others. This allows you to have a pristine startup snapshot of the machine's capabilities. Anything you do to the machine from here on out - add software / hardware, tweak registry settings, run applications, and so on - will change your machine's performance, so you need to know what affects your changes have to the machine.

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.

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

Update the drivers! This is something a surprising number of people do not do, particularly with new workstations. Immediately check the 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.

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.

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 the Building Design Suite 2015 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.

Matt's Workstation Buying Guide, 2014 Edition

In this section my goal is to build three separate classes of machines: a solid, entry level workstation based on the Haswell desktop processor, a Haswell-E based workstation, and a Xeon E5 based system.

For each system class I was going to strive to build two machines: one from components priced from Newegg.com, the other a Dell Precision Workstation, compare prices and declare a winner. However, it became apparent that while it's easy to find desktop based parts and build equivalent entry-level workstations, it's much more difficult to build out a Xeon based workstation that would be considered cost effective from parts. You also come up against configuration limitations with purchasing vendor workstations, as I explained previously.

Furthermore, in the quest for the lowest price, I found that in some cases purchasing Dell equipment from someone other than Dell reaped solid rewards. Those items are indicated below.

Due to these limitation I decided to build some hybrid machines, purchasing a bare bones Dell Precision workstation with the right CPU, and supplant that with components I can get from Newegg.com.

What separates one class of workstation from another is the variation in CPU, motherboard, amount of RAM, and video card. I'll use the same components for the rest of the builds: RAM, storage, mouse, keyboard, etc.

Terms:

- I'm pricing the Dell workstation anonymously from the online site. If I requested actual quotes from human beings, even salespeople, the price would probably be lower.
- I'm including dual 27" high resolution IPS panel LCD screens in all builds.
- No MS Office, productivity software, security software, or specific vendor tweaks are included in any builds.
- For the Dell systems all DVD resource disks for the OS and drivers are included in the price.
- For Dell systems I will purchase the monitors, mouse, and keyboard from them where it makes sense. Sometimes (depending on the day, hour, minute) pricing advantage may swing either way.
- 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 780 Ti. While this is last year's Kepler model, it has a full complement of 2,880 CUDA cores, the most of any graphics card. Being last year's model, it's also a lot less expensive than this year's GTX 980 and performs better in iRay.

The Entry Level BIM Workstation: Dell Precision

First up is a Dell Precision T1700 Workstation with a Haswell Core i7-4790 processor.

Entry Level BIM Workstation: Dell T1700 Precision Workstation, Mini Tower		
Component	Item	Price
Processor	4th Generation Intel Core i7-4790 @ 3.6GHz	$1,862.24
Memory	16GB (2x8GB) 1600MHz DDR3 Non-ECC	
Graphics	Integrated Graphics Processor on CPU	
Storage	256GB 2.5" Solid State Drive (model unknown)	
	1TB 3.5" 7200 RPM Hard Drive (model unknown)	
	16x DVD+/-RW SATA (model unknown)	
Power supply	365W 90% Efficient Power Supply	
Keyboard	Dell KB-522 Wired Business USB Multimedia Keyboard	
OS	Windows 8.1 Professional, 64-bit	
Warranty	3 Year ProSupport + 3 Year NBD Limited Onsite Service after Remote Diagnosis	
Mouse	Logitech MX Performance Mouse	$99.99
Monitor	Dell UltraSharp U2713HM 27-inch Widescreen Flat Panel	$639.99
	Dell UltraSharp U2713HM 27-inch Widescreen Flat Panel	$639.99
Dell Subtotal		**$3,242.21**
Additional items purchased separately from NewEgg.com		
Video Card	EVGA GeForce GTX 780 Ti w/3GB	$439.99
Newegg Subtotal		$439.99
System Total		**$3,682.20**

The Entry Level BIM Workstation: Newegg

Next up is a comparable home-brewed BIY Core i7-4790K system completely from Newegg

Entry Level BIM Workstation: Newegg		
Component	Item	Price
Case	Corsair Obsidian 750D ATX Full Tower	$134.99
Processor	4th Generation Intel Core i7-4790K @ 3.5GHz	$299.99
Motherboard	ASUS Z97 Pro LGA 1150 Motherboard	$152.99
Memory	Kingston HyperX Black 16GB (2x8GB) DDR3-1600 Non-ECC	$154.99
Graphics	EVGA GeForce GTX 780 w/3GB	$439.99
Storage	Samsung 850 Pro 256GB 2.5" Solid State Drive	$159.99
	Western Digital Black 1TB 3.5” 7200 RPM 6GB/s Hard Drive	$69.99
	Samsung 24x DVD +/-RW SATA DVD Burner	$13.99
Power supply	Corsair AX Series AX 860 860W 80+ Platinum PSU	$169.99
Mouse	Logitech MX Performance Mouse	$74.99
Keyboard	Corsair Gaming K70 w/Cherry MX Red switches	$169.99
OS	Windows 8.1 Professional, 64-bit, OEM	$99.99
Monitor	Dell UltraSharp U2713HM 27-inch Widescreen Flat Panel	$599.99
	Dell UltraSharp U2713HM 27-inch Widescreen Flat Panel	$599.99
System Total		**$3,141.86**

The Entry Level BIM Workstation: Build Analysis

The term “entry level” is a little bit of a misnomer, because this system will perform adequately well for the vast majority of the tasks you will likely do under the Building Design Suite. I attempted to build the machine as powerfully as I could, given the constraints of Dell’s configuration allowances. Note: The same Dell T1700 Precision Workstation using the Xeon E2-1270 v3 will be almost identical in price.

As you can see, almost the entire system can be bought directly from Dell. However, I wasn’t willing to accept the “workstation” class video cards Dell offered, so I specified the machine with the built in IGP and picked up the GTX 780 Ti from Newegg.

Overall the Newegg.com system was about $540 less than the Dell for the same basic configuration. Actually, if I purchase the U2713HM monitors from Newegg instead of Dell I would save another $80, making the real difference around $430. In addition to being less expensive, the Newegg system includes at least the same or higher quality components. Highlights include a highly rated SSD (we do not know what would ship with the Dell), a very nice and roomy case, a much larger and exceptionally good power supply, and a very high end keyboard.

Downsides are what you would expect of home-built systems: No on-site service warranties, higher propensity for the need to RMA back a defective component, and the time it takes to build it and install the OS.

The Ivy Bridge-E Workstation: Newegg

Stepping up to Haswell-E and 6 CPU cores, we have the following Newegg build. Dell does not offer any workstations with the Haswell-E processor, so this is strictly a home brew.

Ivy Bridge-E Workstation from Newegg.com		
Component	Item	Price
Case	Corsair Obsidian 750D	$134.99
Processor	Intel Core i7-5930K Haswell-E 6-Core 3.5GHz LGA 2011-v3 CPU	$579.99
CPU Cooler	Corsair Hydro Series H75 Closed Water Loop CPU Cooler	$69.99
Motherboard	ASUS X99-Delux LGA 2011-v3 ATX Motherboard	$384.99
Memory	Kingston HyperX Black 32GB (4x8GB) DDR3-1600 Non-ECC	$309.98
Graphics	EVGA GeForce GTX 780 Ti w/3GB	$439.99
Storage	Samsung 850 Pro 256GB 2.5" Solid State Drive	$159.99
	Western Digital Black 1TB 3.5” 7200 RPM 6GB/s Hard Drive	$69.99
	Samsung 24x DVD +/-RW SATA DVD Burner	$13.99
Power supply	Corsair AX Series AX 860 860W 80+ Platinum PSU	$169.99
Mouse	Logitech MX Performance Mouse	$74.99
Keyboard	Corsair Gaming K70 w/Cherry MX Red switches	$169.99
OS	Windows 8.1 Professional, 64-bit, OEM	$99.99
Monitor	Dell UltraSharp U2713HM 27-inch Widescreen Flat Panel	$599.99
	Dell UltraSharp U2713HM 27-inch Widescreen Flat Panel	$599.99
System Total		**$3,878.85**

The Ivy Bridge-E Workstation: Build Analysis

For this system I decided to step things up quite a few notches in the most important areas. This is a solid BIM workstation, about $737 more than the entry level system (or, about $200 more than a Dell entry level T1700 workstation, if you prefer), and certainly geared to someone doing more with the Building Design Suite than just AutoCAD and a little Revit.

Specific improvements over the T1700 / entry level workstation are:

1. The CPU was upgraded to the 6-core Intel Haswell-E i7-5930K running at 3.5GHz. If you do a lot of 3ds Max rendering or work on very large models in Revit, you want more than 4 cores running at a good speed. The Haswell-E is about 20% faster than the Devil’s Canyon chip. For $500 more you could upgrade to an 8-core i7-5960X.
2. The ASUS X99-Deluxe motherboard is arguably one of the best values in a Haswell-E motherboard, equipped with 8 RAM slots supporting memory to 64GB, 8 SATA 6GB/s ports, 3 x PCIe 3.0 x16, 2 x PCIe 3.0 x16 (x8 mode), and 1 PCIe 2.0 x4 slots, allowing for multi-GPU configurations of x16; x16/x8; x8/x8/x8. Perfect for mid-Level multi-GPU rendering.
3. System RAM was upped to 32GB (4x8GB) DDR3-1600. Two reasons: (a) We need to fill at least four RAM slots, and (b) Why not?

The High-End Xeon E5 Dell Workstation

This is a straight up Dell Precision T5810 Workstation base system that we will optimize with parts from Newegg.com to make this as least expensive as possible. We purchase it with an NVS graphics card because the system does not ship without something for a graphics card.

Component	Item	Price
Processor	Intel® Xeon® Processor E5-1650 v3, 6-Core, 3.5 GHz, 15 MB	$3,104.50
Memory	32GB (4x8GB) DDR4-2133 RAM	
Graphics	NVIDIA NVS 310	
Storage	256GB 2.5" Solid State Drive (model unknown)	
	1TB 3.5” 7200 RPM Hard Drive (model unknown)	
	16x DVD+/-RW SATA (model unknown)	
Power supply / Chassis	Dell Precision T5810 685W TPM Chassis	
Keyboard	Dell KB-522 Wired Business USB Multimedia Keyboard	
OS	Windows 8.1 Professional, 64-bit	
Warranty	3 Year ProSupport Service with 3 Year NBD Onsite Service after Remote Diagnosis	
Subtotal		**$3,104.50**
	From Newegg.com	
Video Card	EVGA GeForce GTX 780 Ti w/3GB	$439.99
Mouse	Logitech MX Performance Mouse	$74.99
Monitor	Dell UltraSharp U2713HM 27-inch Widescreen Flat Panel	$599.99
	Dell UltraSharp U2713HM 27-inch Widescreen Flat Panel	$599.99
Subtotal		$1,714.96
System Total		**$4,819.46**

To save even more money, we could opt to purchase the system with only 8GB of RAM and a 1TB hard disk and purchase the 32GB RAM and SSD from Newegg. However, that means that we need to install the RAM and copy the system drive to the SSD and format out the hard disk, which may negate any possible savings from simply buying these components from Dell originally.

Links

Industry Pressures

Setting up an Amazon EC3 render farm
http://area.autodesk.com/blogs/cory/setting-up-an-amazon-ec2-render-farm-with-backburner

Revit Specific

Revit Model Performance Technical Note 2014:
http://images.autodesk.com/adsk/files/autodesk_revit_2014_model_performance_technical_note.pdf

Revit on a Mac:
http://images.autodesk.com/adsk/files/autodesk_revit_architecture__mep__structure_mac_faq_final.pdf

Processors:

Intel Tick-Tock Model
http://en.wikipedia.org/wiki/Intel_Tick-Tock
http://www.intel.com/content/www/us/en/silicon-innovations/intel-tick-tock-model-general.html

Introduction to Haswell
http://www.techspot.com/review/679-intel-haswell-core-i7-4770k/

Intel's Comparative Database:
http://ark.intel.com/

Haswell (4th Generation) CPUs Compared:
http://ark.intel.com/compare/80807,80806,77656,75123,75122

Haswell Devil's Canyon review
http://www.anandtech.com/show/8227/devils-canyon-review-intel-core-i7-4790k-and-i5-4690k

Haswell-Based Xeons E3-1200 Series Review
http://www.tomshardware.com/reviews/xeon-e3-1275-v3-haswell-cpu,3590.html

Transactional Synchronization (TSX Instructions) in Haswell:
http://software.intel.com/en-us/blogs/2012/02/07/transactional-synchronization-in-haswell

Inside Haswell:
http://www.anandtech.com/show/6262/intels-haswell-20x-lower-platform-idle-power-than-sandy-bridge-2x-gpu-performance-of-ivy-bridge

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/syscert?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 Review: http://www.anandtech.com/show/8526/nvidia-geforce-gtx-980-review

NVIDIA GTX 780 Ti review: http://www.tomshardware.com/reviews/geforce-gtx-780-ti-review-benchmarks,3663.html

BOXX Blogs: GeForce GTX Rendering Benchmarks and Comparisons
http://blog.boxxtech.com/2014/11/17/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

BOXX Blogs: GTX 780 GPU Rendering Benchmark with Octane Render
http://boxxblogs.blogspot.com/2013/06/gtx-780-gpu-rendering-benchmark-with.html

iRay Rendering Benchmarks: http://www.migenius.com/products/nvidia-iray/iray-benchmarks

OpenCL and CUDA are Go: GeForce GTX Titan Tested in Pro Apps
http://www.tomshardware.com/reviews/geforce-gtx-titan-opencl-cuda-workstation,3474.html

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

Kepler GK110 White Paper: http://www.NVIDIA.com/content/PDF/kepler/NVIDIA-Kepler-GK110-Architecture-Whitepaper.pdf

SIGGRAPH 2012 - NVIDIA talk on Advanced Rendering and GPU Ray Tracing:
http://NVIDIA.fullviewmedia.com/siggraph2012/ondemand/SS107.html

IRay Performance Tips (from 2011)
http://irayrender.com/fileadmin/filemount/editor/PDF/iray_Performance_Tips_100511.pdf

GTC on demand - GPU Ray Tracing presentation
http://www.gputechconf.com/gtcnew/on-demand-gtc.php

NVIDIA mobile GPU comparison sheet:
http://www.nvidia.com/object/quadro-for-mobile-workstations.html

Memory:

Sandy Bridge Memory Scaling: Choosing the Best DDR3
http://www.anandtech.com/show/4503/sandy-bridge-memory-scaling-choosing-the-best-ddr3/8

Peripherals:

Cherry MX Switches:
http://www.keyboardco.com/blog/index.php/2012/12/an-introduction-to-cherry-mx-mechanical-switches/

System Builds:

Haswell / GTX 780 PC Build
http://www.maximumpc.com/build_a_780_haswell_gaming_PC