# New article: Intel's Near-Threshold Computing

Article: Intel's Near-Threshold Voltage Computing and Applications

By: **Giovanni Lostumbo** (giovanni.lostumbo.delete@this.gmail.com), December 26, 2024 2:26 pm Room: Moderated Discussions

David Kanter (dkanter.delete@this.realworldtech.com) on September 18, 2012 11:26 am wrote:
> Our latest article has just gone online:
>
> "Near-threshold voltage computing extends the voltage scaling associated with Moore's Law and
> dramatically improves power and energy efficiency. The technology is superb for throughput,
> at the cost of latency, and best suited to Intel's products for HPC and mobile graphics."
>
> This article is a detailed look at the circuit techniques required to operate down to 0.3-0.5V using standard
> CMOS, based on several papers published by Intel at ISSCC. The implications of the technology are explored,
> including a discussion of which markets and SoC components are most amenable to near-threshold computing.
>
> http://www.realworldtech.com/near-threshold-voltage/
>
> As always, feedback and comments welcome.
>
>
> David

"Moreover, the optimal operating point for NTV designs tends to be quite low. The 32nm Pentium core increased efficiency by about 5×, by running at slightly under 100MHz. The maximum frequency was 915MHz, so the absolute performance decreased by about an order of magnitude. That is a tremendous sacrifice to achieve energy efficiency; one that may not be feasible for many applications."

"The trade-offs associated with near-threshold voltage techniques strongly suggest certain applications. General purpose CPUs for client systems are unlikely to benefit from NTV. While energy efficiency is important, sacrificing frequency is inconsistent with the overall design targets."

This assessment was wrong at the time and is even more wrong today.

The HPC and exascale applications do get the most energy savings, but the "tremendous sacrifice" for many applications does not preclude all applications. The ntv processor is ideal for energy harvesting systems like ONiO and e-peas powered microcontrollers. Furthermore, ntv can run Linux <2.4 and Windows 95/with very low RAM requirements. Modernizing these systems while retaining microkernel or modular systems could offload additional memory requirements (and thus, lower power) to cold storage.

This conclusion in 2012 may have sounded practical at the time, as HPC and the server market would, comprise a larger potential revenue source for chips utilizing NTV. However, as time went on, sub-threshold companies such as Ambiq Micro have demonstrated 48-250MHz on 12-40nm TSMC ULL chips with 4MB of RAM. This exceeds the speeds of Pentium IIs despite having far less transistors on the Cortex M4 and exceeds Windows 3.1 memory requirements.

The developing world, in regions without electricity, could benefit from having a solar-powered phone and laptop that can run Microsoft Encarta or a web browser from the 1990s. Citizen Band radio (CB) is a public radio channel that could be analogized to unencrypted internet. Systems with low resources did not use encryption. AES-NI wasn't x86 integrated hardware based until 2010. Do new laptops need TPM 2.0 to be "secure enough"? Will there ever be an off-ramp for edu-tech that isn't based on Giga or Teraflops minimums? Who makes these design decisions? People who want their children to have better computers than they had when they grew up? I learned computers with a Packard Bell Legend 3540 in 1996, which was a 100MHz Pentium with 8MB RAM. I had MS Encarta 95' and didn't have dial-up for nearly 2 years until 1998. What good is a Chromebook in the 3rd grade when it doesn't even have any offline apps? In the 3rd grade, my school was still using Apple IIs for Oregon Trail and Number Munchers on 5 1/4" floppies that students manually loaded. I didn't use System 7 Macs until the 7th grade. There is a enduring utility in the WIMP (Windows, Icons, Menus, and Pointers) interface, something that capacitive touch screens don't teach, kind of like getting rid of cursive writing instruction. While Windows 95 could run on 8MB of RAM (4MB too), the earliest WIMPs such as the Apple Lisa in 1983 used less than 128K of RAM. A microcontroller today has more RAM than that, but it's a faux paux to call them computers today?

Encryption is important, of course, but proof of concept buys hearts and minds first. The Incompatible Timesharing System at MIT's Artificial Intelligence laboratory in 1967 initially had no passwords. What better way to deny students a educational moment by completely omitting an aspect of computer history when they first logon? Guest access to systems is a design feature that too few systems today offer, and not because of security concerns, but because of ideological coddling of the American spirit. For starters, write access can be disabled if one believed a guest could tamper with such a system. Security upgrades are a never-ending goal post movement. Likewise, books do not have locks on them, but digital encyclopedias aren't commodified the same way, because of dependence on the cloud. Even an out-of-date encyclopedia is correct twice a day.

Wireless radios like Digital Radio Mondiale (DRM) is a system that provides low-cost, low power radio to many parts of the world that previously required more power-intensive radio receivers. Laptops need the same disruptive technology- and cpus designed for solar power in mind would help bridge that gap. Intel is at its 15th generation of i processors since Nehalem and yet can't even make a 2nd generation scientific calculator that can display Windows 95. Why did they didn't they market the Claremont?

Who is to say that sacrificing frequency is inconsistent with design targets when not everyone in the world has access to dial-up internet speeds? I think some, not anyone in an industrialized world that has become accustomed to high speed internet would think about for needing themselves, and would feel like a downgrade, but some one in the developing world would prefer to have something over nothing. That's cultural relativism at its finest. I learned computers on 100Mhz, yet the OLPC was 450MHz and it was called "underpowered." Windows 95 was written in C. RISC OS was written in Assembly language. Both are closer to metal than Windows XP+ and Sugar OS (Python) was. That's why it was "Slow".

These are the design decisions that should be factored into a educational laptop- that provides the minimum code needed for basic interactive computing:
https://github.com/kragen/dernocua/blob/master/text/energy-autonomous-computing.md

EDITOR'S PICKS

**A Historical Look at the VAX: Economics**

By John Mashey

**Why Apple Won't ARM the M**

By David Kanter

**Intel's Sandy Bridge Microarc**

By David Kanter

RWT ON TWITTER

- RT @TheKanter: Ashlee is a writer! You won't be disappoi about 1 year ago
- Someone loves our forums!!! https://t.co/4flSfwtEQr abo
- @CRN_WMid I totally agree midwives, but did you know t account of a tech... https://t about 2 years ago in reply t

Follow @RWT

Personally, I tried using RISC OS only in the past few years- its UI seems a bit awkward, the mouse pointer isn't as intuitive as Microsoft's Windows 95 and up, and the austere windows are as barebone as Windows 3.1.

But if 450MHz isn't fast enough for a GUI, it's not the CPU that's slow, it's either the OS, the window manager, or the apps that are bloated/slowing down the PC, or possibly the user that doesn't know how to use the system, also possibly due to bad UI design. The 32-bit color era of the mid 90s was the most code dense time for computers. They communicated as much information visually as they did in operations.

According to a former Intel engineer, an x86 license can cost $10 billion.
If Nvidia, Mediatek, and Qualcomm, which can afford it, have no interest in developing a solar powered laptop because it's not immediately profitable- with relatively slow performance, no one else might develop it. Thus an architecture that is out of patent, such as the SuperH instruction set, which expired in 2015, could run linux and be developed at Skywater's 90nm- FD-SOI, or Global Foundry's FDX 22nm FD-SOI. But it seems like the OLPC and other developers would rather see their project fail before it could even implement near-threshold voltage since they received so much criticism. So many assumptions made on infrastructure- expecting third world countries to have reliable utility access and assuming a NGO could reliably develop an electric grid and fiber-optic internet when rooftop solar panels are a very visible target for theft and resale.

The NASA scientists that gleefully await 2 days for roundtrip communications the Voyager 1 on its backup S-band might not balk at the idea of waiting a few extra minutes, or even hours, to receive kilobytes of data from remote, non-Western regions of the world. In fact, I think fulfilling an original goal of the internet, to improve connectivity to other parts of the world, would be made possible by specific developing ntv for general purpose computing. I admit I do not know if that was an original goal.

State of the art chip designers should allow the rest of the world to experience the optimism that Westerners had in the 1990s before suggesting other new technologies in artificial intelligence are the only foregone conclusion. With all the retro Nintendo and Sega game systems sold in miniature forms, Intel could have released a miniature Pentium 32nm, 22nm, or 12nm FF, integrated a passive matrix display with ultra low power consumption, or even memory in pixel (MIP) display. Casio, Texas Instruments made solar powered calculators in the late 70s and 80s, but curiously Intel never seemed to think their ntv tech could be an interactive educational tool. According to Yahoo Finance, a December 19th article said:

"But company insiders and industry analysts tell Yahoo Finance that Intel's dramatic crash is the result of a slow deterioration spanning more than two decades.

“They had a God complex; they were super arrogant,” a former high-level executive who worked at Intel for more than 20 years told Yahoo Finance. “They felt like they had such a large competitive advantage that they could never do anything wrong.””

"But after the dot-com bust, Intel invested in multiple projects that never materialized or failed to reach their potential. Two former executives told Yahoo Finance that innovative efforts were often killed if they didn't immediately contribute to revenue or risked cannibalizing existing products.

One former high-level executive, who worked within several divisions, said Intel didn't support the team working on low-power Atom chips for mobile phones in the early 2000s. It sold its license for Xscale, then Arm's most advanced architecture for mobile chips, to Marvell (MRVL) in 2006."

https://finance.yahoo.com/news/how-innovation-died-at-intel-americas-only-leading-edge-chip-manufacturer-faces-an-uncertain-future-and-lawsuits-130018997.html

That includes Medfield (18mW idle power https://www.anandtech.com/show/5365/intels-medfield-atom-z2460-arrive-for-smartphones), the Quark, and the Claremont that they were based off of. There could have been a 2mW Pentium for sale by 2015, and could be powered by a lithium ion capacitor with a 90mAh battery life.

I contacted Intel's Foundry in April of 2021, asking them if they would allow licensing of the Quark for solar powered laptops. No response. I contacted them last week. Still no response. Intel hasn't changed at all.

In fact, it would have solved one of the problems of the OLPC's battery life, and the Intel ClassMate PC could have used it with 8-16 MB of RAM, provided it wasn't trying to run Windows 2000 or above. It's kind of like someone trying to sell me a Nintendo Switch because it runs 1GHz/and not 1.79mHz like the NES. It's not the nostalgia, it's the learning experience of discovering more than one form of human-computer interaction. And it's the quality of the information that counts, not the bytes. Like that MasterCard commercial, "there are some things money can't buy, but for everything else, there's MasterCard." 3D is optional. 2D GUIs should be the first exit ramp of technological deconvergence from all the "standard" features expected of an OS/system today.

These opinions are directed at the technology industry at large. When Steve Jobs saw the Alto systems at Xerox PARC, he immediately knew that was the future. But the people selling AI aren't selling the whole picture. Because I know the universal potential of NTV. "You’ve got to start with the customer experience and work backwards to the technology." That's why Jobs, who was not an engineer, could define a product that people wanted, better than the design goals/agenda of data-mining chip architecture. "If something is free, you are the product." And even if something isn't free, a user may still be the product. Koomey's law has reached the point where near and sub-threshold voltage chips can accomplish 1980s, 1990s, and now 2000s-level CPU performance in less than 5 milliwatts, and the only thing preventing them from running on solar power is the extra circuitry that is needed for AES, Intel Management Engine, TPM, and Windows 11 screen recording, because the overhead is predominantly user-tracking and not essential to WIMP navigation. So while a Pentium II, which had 8 million transistors, might run on solar power, the 20+ million transistors needed to track every cursor movement, click and system log will ensure solar powerable laptops aren't made possible, if ever, since it would require 18NA transistors to consume sub 5 milliwatt system TDP with all that extra overhead.

Paired with an LTE Modem, it would likely need to consume far more power to upload all that telemetry, since so many apps are constantly updated from the Google Play store with background data usage that would saturate a 2G connection. Thanks to
Wirth's Law, "What Intel Giveth, Microsoft (and Google) taketh away." But it's Intel that now also attempted to destroy their own product line up with X86S, ensuring legacy devices remain an obscure and downplayed value. Fortunately, they seemed to have backed off of that initiative before solar chips could ever have their day in the market. Perhaps their marketing strategy is, "If we try to wean enough people off 32-bit, maybe the goal posts for solar power will be made even more difficult for idealists. Since 64 bit architecture requires 10-30% more memory, and memory uses around 8x the number of transistors of logic, then a 100MHz Pentium with 8MB of RAM now needs 9-12MB of RAM to complete the same operations. How many 64 bit CPU/microcontrollers under 100MHz are there? Does every application need 64 bit? Or is it something that "might be nice to have"? And at what cost? Sacrificing portability and energy independence? I doubt it's worth being tethered to a USB cable for topping up the charge for another decade to use a word processor, play Solitare, and run IRC. In the early 2000s, the demoscene had a 64K challenge, which was making 3D demos under 64K https://en.wikipedia.org/wiki/Demoscene#64K_intro). Maybe it's time for a 16MB challenge- make more software that fits under 16MB files, and 272MB filesystems.
https://en.wikipedia.org/wiki/Cramfs And a 4GB challenge for 32-bit OSes. ext for linux had a 2GB file size limit. Windows 95 took 75MB of disk space. An entire OS could run in less than half of a liveCD and boot from ROM and provide ample functionality for local and remote desktop use. A minimum viable product today is defined more by profitability than utility, but this type of laptop is a mythical unicorn:
https://medium.com/this-should-exist/prose-a-distraction-free-e-ink-laptop-for-thinkers-writers-4182a62d63b2

All the linux developers in the 90s built linux on a 32 bit system, sometimes using 16 or even 8 bit computers. With so much messaging

encouraging new students to "learn to code," it would be far more useful to make the means of computing accessible (to hundreds of millions of more users) without the imperative of learning to code, and I would go so far to say that ntv is probably the 3rd most significant development in hardware after the GUI and the 300 baud modem. Information asymmetry is due to the information bottlenecks and filter bubbles not just in TPOTs, but caused by power consumption requirements to run a mobile device in the tens or hundreds of milliwatts. When the OLPC was dropped off in some villages, little consideration was made for battery life. 2 Watts is not instant on like a solar calculator. Importing not just the client software, but the power plant on the device- the infrastructure needed to operate as a node in a mesh network, or, at the very least, a self-powered thin-client, allows the user to rely as little as possible on the local externalities that may be scarce and/or expensive. Brownouts, political instability, price-gouging, etc. It also costs far less to deploy a terrestrial cell tower than the 24 satellites needed for GPS and worldwide navigation, and wired internet is sometimes prohibitively expensive to lay copper down (although there are some auto-boring machines that can lay cables without the need to dig trenches). And that is why ntv would be the most impactful technology addition to computing since baked silicon.

Lastly, consider the double standard used in processing radioactive uranium. https://world-nuclear.org/information-library/nuclear-fuel-cycle/conversion-enrichment-and-fabrication/uranium-enrichment Governments around the world consider even a 5% enrichment of uranium practical for power reactors, but not utilizing 95% of the performance of a Pentium is a "tremendous sacrifice"? Computers to Steve Jobs were like a bicycle, not an aircraft carrier: https://youtu.be/KrPsyr8Ig6M?t=857

IN 1992, at CERN: A disruptive victory of bottom-up over top-down - Ben Segal
"At CERN we wanted to solve a technical problem (sharing data between many different sorts of computers"
• It worked well but we were not supposed to go beyond "testing":
we were forbidden to connect machines outside CERN"

"We reject: kings, presidents, and voting. We believe in:
rough consensus and running code." - Dave Clark, 24th IETF meeting, 1992
https://indico.cern.ch/event/1331906/contributions/5606846/attachments/2743709/4773509/TCPIP_CS40.pdf (p.7/14)

There is a still a de-facto barrier to entry in the digital divide- that is electricity access. Removing the need for wired power is just as important, if not more important than the interoperable TCP/IP protocols. Freedom from a Grand Old Utility is priceless. No one would expect a computer to rely on gas powered generators, diesel engines, coal factories just to send a text message. Thus solar power could address at the very least, the low-hanging fruit of offline basic input/output command line interfaces, interactive programming with a completely separate power supply for energy-lite applications, things that do not require a 3000mWh battery. In fact, Lenovo had a dual screen e-ink laptop for energy "saving" displays, and Hisense has a dual screen cell phone (A6L) with e-paper on one side and color LCD on the other side. https://www.youtube.com/watch?v=opN9gtlz4zQ While a phone that has a low power mode might be available on most modern Android (Super Battery Saving Modes), there's nothing wrong with adding a little discrete redundancy in both power saving and battery charging systems.

NTV is the bottom up client hardware complement to TCP/IP, because at no layer should internet protocols be hindered by the friction of cruft and power. The speed bumps that exist to slow internet thought are more to prevent internet rage. But today's infrastructure are still held by gatekeepers that assume someone will fix the internet infrastructure somewhere, when building the tools for users to help themselves connect to the interent should be a top priority, above new architectures and languages that further fragment the software ecosystem. If Intel or AMD developed solar powered laptops, you could have millions more Rust developers or X86S developers who do not need reliable electricity mains to be running. A Raspberry Pi costs $120 in Brazil due to protectionism, but even if it were free, each future coder would need a monitor, a keyboard, a USB charger to learn coding. Rather than try to accomplish everything on a single board computer, the technology that allows a microprocessor to run on 1/10th of the RAM of a Raspberry Pi Zero would still allow a user to learn coding on 32MB of RAM if it could run an early linux like in this video: https://github.com/hatonthecat/linux_distro_tests?tab=readme-ov-file#virtualbox

With so much emphasis on microcontrollers and small Arduinos, it emphasizes a high technical proficiency as a barrier to entry for casual users. I might learn how to use a UART/Serial firmware, but it doesn't make me a better coder. With programming requiring a highly logical skillset, more accessible computing could instead emphasize a balance between visual coding and command-line interfaces, and single board computers don't need 32MB of video memory like in the Raspberry Pi. My Packard Bell had 1MB of video RAM. and yet it was sufficiently colorful and rich in text to learn how to navigate a friendly user-interface that has remained largely unchanged (and for good reason), beyond optionally accelerated and smooth window animations/transitions. With the 30 year anniversary of Windows 95 nearing, perhaps Microsoft and Intel can open source the OS and P54C that led to a generation of users who could actually use computers as multimedia PCs and not just command-line interfaces.

But I admit, I never had to learn slide rule in school. This was before portable calculators. Perhaps if I had to learn more manual techniques, I might be more deferring to terminal-centric programming. But chip design faces the same groupthink as the linux ecosystem. There is no year of the linux desktop because it fails to reach a large market outside of Android. If chip designers and kernel engineers build a chip designed for portable solar panels (the size of a credit card or notecard, not the size of a backpack), the linux community would double within years. Like the movie Field of Dreams, "If you build it..."

< Previous Post in Thread

Next Post in Thread >