The Occupancy Gap: Why the Same GPU Could Mine BTX ~2.5x Faster (July 2026)

A BTX rig often left a third of its GPU idle. We benchmarked the gap on a stock RTX 3060, in nonces per second and watts, and explain the one lever that closed it. Measured in July 2026, before BTX changed its proof of work, so read the rates here as a record of the old engines rather than a current benchmark.

JULY 6, 2026 · 6 MIN READ · EASYBTX.COM/RESEARCH/BTX-GPU-OCCUPANCY

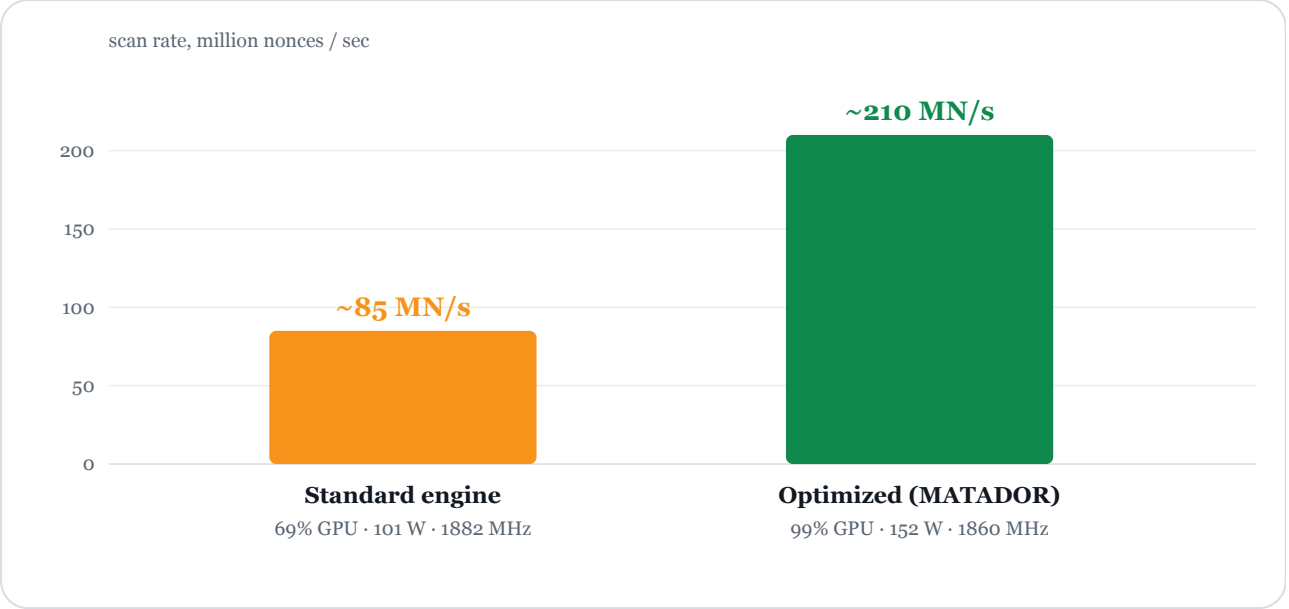
ABSTRACT

Most BTX miners ran their graphics card at about two-thirds of its capacity without knowing it. We measured the gap on a real RTX 3060, same card and same clocks, and explain what actually closed it. Measured July 2026, before the MatMul v4.7 fork retired the nonces-per-second unit.

There is a comfortable assumption behind most GPU mining: that when your rig is mining, your graphics card is running flat out. For BTX, that assumption is usually wrong. On a typical setup the card sits at roughly two-thirds of its real capacity, and the missing third is not lost to heat or bad luck, it is simply never used. We call it the occupancy gap, and it is the single largest, least-understood lever in BTX mining performance.

We measured it directly. One stock RTX 3060, nothing overclocked, mining live to a real pool, run twice: once on a standard BTX mining engine, once on a heavily-optimized one. Same silicon, same drivers, same clocks. Only the kernel, the code that actually drives the GPU, changed.

What we measured



Live scan rate on one stock RTX 3060, standard engine versus an occupancy-optimized engine. Same card, same clocks. The difference is almost entirely how fully the GPU is used.

The standard engine held the card at about **69% utilization** (peaking in the high 70s) and produced roughly **85 million nonces per second** while drawing about **101 watts**. The optimized engine pinned the same card at a flat **99% utilization**, produced about **205 to 214 million nonces per second**, and drew about **152 watts**. Zero rejected shares on either. The core clock barely moved: 1882 MHz versus 1860 MHz.

That is a **~2.4 to 2.5x** improvement in useful work, on identical hardware, with no overclock. The entire difference is the occupancy gap being closed.

Why the gap exists

BTX proof-of-work is not a simple hash loop. Each nonce runs through a matrix-multiplication stage and a hash-scan stage, streaming data through the GPU's memory. A straightforward implementation of that pipeline spends a lot of its time waiting: the GPU's many streaming multiprocessors, the units that do the actual math, stall while data moves in and out of memory. Every cycle an SM waits is a cycle of the card you paid for, doing nothing.

A well-engineered kernel attacks exactly that. By restructuring how work is scheduled and how data is kept on-chip, it keeps the SMs fed so they rarely stall. The card stops idling at 70% and holds 99%. No new hardware, no higher clock, just a design that refuses to leave the silicon waiting.

This is why the number on the box, or the raw specification of your card, tells you surprisingly little about how fast it will mine BTX. The kernel decides how much of that card you actually get.

The efficiency most miners miss

The instinct is to see 152 watts versus 101 and conclude the faster engine is simply burning more power. Look closer. It draws about **1.5x the power** and delivers about **2.5x the work**, roughly **1.65x more nonces per watt**. A GPU has a large fixed cost just to be powered on and clocked; spreading that cost across far more useful work is what makes a fuller card a more efficient one. On a power meter, closing the occupancy gap is a win, not a cost.

The honest trade-off

There is a catch worth stating plainly. The fastest BTX engines available today are **closed, third-party binaries**. The leading one takes a **1% developer fee** and, as the numbers show, runs your card hotter and hungrier. It mines only to your own payout address, but it is not open and not free.

That is a real choice, and it should be a choice, not a default forced on anyone. easyBTX ships its own **free, open engine** by default, which most people should keep. As of **v0.10.0**, it also offers the faster closed engine as a clearly-labelled opt-in, **MATADOR mode**, with the 1% fee, the extra power, and the closed nature disclosed up front. Flip it on for maximum speed; flip it off to return to the open engine. The lever is yours, and nothing about it is hidden.

Why this matters beyond one rig

Zoom out from a single card to a whole fleet. When honest miners get 2.5x the real work out of the same hardware, more of the network's proof-of-work is done efficiently by people who actually hold and use BTX, instead of being quietly left on the table by cards running at two-thirds idle. A fleet of well-utilized, honestly-run rigs is a more secure and more decentralized chain than the same hardware half-asleep.

That is the quiet thesis under all of this. The gap between 70% and 100% of your GPU is not just your gap. Summed across every miner who closes it, it is hashpower that stays in the hands of the community running the chain. Combine stronger, and there is more BTX for everyone doing the work.

Postscript added 22 August 2026
Everything above was measured on 6 July 2026, five weeks before BTX changed its proof of work. The measurement stands as a record of the pre-fork engines. The unit it is written in does not survive the fork, so read it as history, not as a current benchmark.

Nonces per second is an obsolete unit. Before the fork, one mining attempt was one nonce, which is why rates were quoted in millions of nonces per second. Since 10 August 2026 one attempt is a single deterministic inference episode, and easyBTX 0.21.1 changed the readout to episodes per minute. In the old units, a perfectly healthy machine now displays a rate of zero. Any MN/s figure in this article belongs to the old work function and cannot be compared with anything measured after block 185,000.

What an episode costs on the same class of card. Post-fork we measured an RTX 3060 at roughly 5.3 seconds per episode under the MATADOR engine, about 11.3 episodes per minute, and roughly 7.96 seconds per episode under btx-rc-miner, about 7.5 episodes per minute. Stated plainly so nobody has to guess: the project's own published reference measurement for one episode is roughly 30 seconds on reference hardware, and we have not established why our pool-side episode times are shorter. We are not going to invent a reconciliation. Until someone shows the working, treat the two as measurements of different things.

Do not carry the 2.5x forward. The occupancy idea generalises, because a stalled kernel leaving streaming multiprocessors idle is a property of GPU scheduling rather than of any one work function. The specific 2.5x number is not general. It was measured on the v3 workload, and we have not re-run the comparison on the new one.

Shares are much rarer now, and that is normal. Because one attempt is thousands of times more expensive, the share counter moves slowly. On a Mac we see roughly one accepted share per 40 episodes on a pool's beginner difficulty, settling to about one per 212 once variable difficulty finds its level. A quiet counter is no longer a symptom.

Frequently asked questions

Why does the same GPU mine BTX at very different speeds?

Because the bottleneck is rarely the raw silicon, it is how well the mining kernel keeps that silicon busy. BTX proof-of-work is a matrix-multiply plus hash-scan pipeline. A naive kernel stalls waiting on memory and leaves many of the GPU's streaming multiprocessors idle, so the card runs at roughly 70% utilization. A kernel that restructures the same work to keep every unit fed can hold the card at 99% and do about 2.5x the nonces per second, at the identical clock speed. It is not overclocking; it is filling the card.

Is this just overclocking the GPU?

No. In our measurements the core clock was essentially the same on both engines (about 1860 to 1882 MHz). The faster engine did not raise the clock; it raised utilization, from the low 70s percent to a flat 99%. Overclocking pushes the same partially-idle card a little harder; closing the occupancy gap fills the idle part. They are independent levers, and occupancy is by far the larger one here.

Does mining faster use proportionally more electricity?

It uses more power, but not proportionally, which is the point. On the RTX 3060 we measured, the faster engine drew about 152 W versus 101 W, roughly 1.5x the power, while producing about 2.5x the work. That is about 1.65x more nonces per watt. A fuller card is a more efficient card, because the fixed overhead of powering the GPU is spread across far more useful work.

What is a nonce per second and how does it relate to hashrate?

BTX uses a matrix-multiplication proof-of-work. Each attempt at a valid block is a nonce, and the scan rate, nonces per second, often shown as MN/s for millions per second, is the GPU-level measure of how fast a rig searches. It is the BTX equivalent of hashrate. The pool credits you in proportion to

the valid shares your scan rate produces, so more nonces per second means a larger slice of each block reward.

Is a faster closed engine safe to use?

It is a trade-off you should make with eyes open. The fastest BTX engines today are closed, third-party binaries, and the leading one takes a 1% developer fee and draws more power and heat. It mines only to your own payout address. easyBTX ships its own free, open engine by default and, as of v0.10.0, offers the faster closed engine as a clearly-disclosed opt-in called MATADOR mode, so the choice, and the 1% fee, is always yours, not hidden.

Will closing the occupancy gap help the BTX network, not just my rig?

Yes, indirectly. When honest miners get more real work out of the same hardware, more of the network's proof-of-work is done efficiently by people who actually hold and use BTX, rather than being left on the table. A fleet of well-utilized, honestly-run rigs is a more secure and more decentralized chain than the same hardware running at two-thirds idle. Efficiency at the edge compounds into security at the center.
