AI Wearables Are Limited by the Radio, Not the Chip

A Technical Analysis of Wi-Fi Front-Ends for Edge AI

By Dr. Ali Arya Fard, CEO / CTO, QuantalRF



QuantalRF

Technical White Paper

Contents

Why Edge AI Changes the Radio's Job	4
What a Cleaner Signal Buys	4
What the Body Does to the Signal	5
Packets Sent Again: The Hidden Battery Cost	6
Energy Per Bit and Wide Channels	7
Why a Training Loop Is the Wrong Tool on a Moving Body	8
QuantalRF 2nd Gen with QuantalRF Antennas	10
3rd Gen: One Radio Outperforms Two	12

Abstract

AI glasses can already see, listen and think. What holds them back is getting what they see to the model. Processors, AI models, cameras and batteries all get better every year. The Wi-Fi front-end, the amplifier that sits between the Wi-Fi chip and the antenna, has not kept pace, and with today's parts it takes about two-thirds of a wearable's power, according to device makers.

The industry's usual answers are more power, more digital correction and more radios. On a device worn on the face, each one adds heat and drains a battery of just 0.3 to 0.9 Wh, and none of them efficiently holds the link's speed once the human body has weakened the signal 400 to 1,600 times. The only answer that scales is a cleaner signal from every milliwatt. That requires rethinking the amplifier itself, and it is what QuantalRF has built.

This white paper explains why, in five steps:

1. A cleaner signal carries more data at the same power.
2. The body takes away most of the signal margin, so cleanliness becomes the deciding factor.
3. A weak or noisy signal forces packets to be sent again, which burns battery and freezes video.
4. A front-end that stays clean on wide channels, up to the full 80 MHz and beyond, moves each bit for less energy, and does not lean on digital correction to stay clean.
5. QuantalRF's Elementum™ 2nd Gen Wi-Fi FEM with QuantalRF DockOn® antennas already beats today's front-ends on every measure we model. Our 3rd Gen Wi-Fi FEM with our antenna sends more data from one radio than an incumbent sends with two, without the extra power of a second radio.

Note: The figures that are marked measured come from bench data on QuantalRF silicon. The figures that are marked modeled come from a link and battery model that is set up to favor the incumbent.



POWER CONSUMPTION

2/3

of the wearable's power
is consumed by the
Wi-Fi Front End.

Why Edge AI Changes the Radio's Job

Phones mostly download. AI wearables mostly upload: the camera and microphones feed a model that runs on the phone or in the cloud, so the device sends a steady stream of video and audio for as long as it is in use.

That turns the transmitter, not the receiver, into the main power load. It also makes the front-end the one part that does not improve with new process nodes. Processors, AI models, cameras and batteries all get more efficient each year. The front-end is analog, so it gains little when chips move to a smaller process.

	Battery Doorbell Camera	AI Glasses
Data it sends	2 Mbps, in short clips	100–500 Mbps, for hours
Battery	about 22 Wh	0.3–0.9 Wh (70 to 25 times less)
Distance to the receiver	10–25 m to the router	under 1 m to a phone, but the body makes it act like 20 m
Signal lost to the body	none	400–1,600 times weaker
Speed setting it needs	the slowest setting works	the fastest settings
Signal over time	steady	changes with every step

A doorbell sends a little data through a wall. AI glasses send a significant amount of data through the body. Today's Wi-Fi front-ends were designed for phones, which have batteries of 15 Wh or more and room to spread heat. They were not designed for the AI glass specification, as noted in the table above.

What a Cleaner Signal Buys

A cleaner transmitted signal lets the link use a faster speed setting at the same power, or the same speed with less power. At each higher speed setting, Wi-Fi packs more bits into each symbol, and each step up needs a cleaner, more accurate signal.

The accuracy is measured as signal error: how far each transmitted symbol lands from where it should be; engineers call this Error Vector Magnitude (EVM). Wi-Fi sets a maximum error for each speed setting:

Speed Setting	Bits Per Symbol	Maximum Signal Error Allowed
MCS7	6	4.5%
MCS9 (AI glass target)	8	2.5%
MCS11	10	1.8%
MCS13 (fastest in Wi-Fi 7)	12	1.3%

The front-end is only one source of error. The Wi-Fi chip, the receiver and the air add the rest. A front-end that uses up most of the error budget forces the whole link down to a slower setting.

Measured: QuantalRF's 2nd Gen part for edge AI, running from 3.3 V on an 80 MHz channel, sends 17 dBm (50 mW) with 0.8% signal error on every channel tested from 5.2 to 6.1 GHz. MCS9 allows 2.5%, so the part leaves about three times more room than the speed setting needs. At 18 dBm it still holds 1.8%.

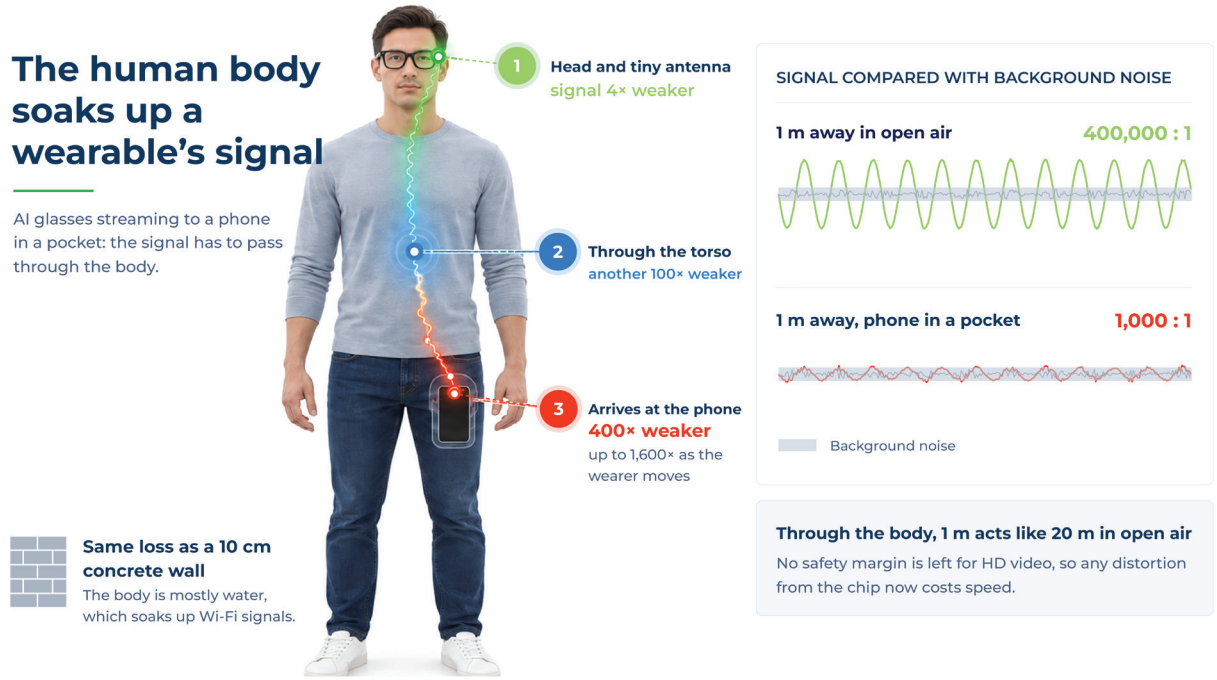
At the error levels incumbent low- and mid-power front-ends publish today, QuantalRF's part sends 6 to 22 times more power from 3.3 V (measured, against published datasheets). That extra clean power is what a wearable spends to win back what the body takes.

What the Body Does to the Signal

When AI glasses talk to a phone in a pocket, the signal crosses the body twice: once at the head, where a small antenna sits against the skin, and again through the torso to the phone. In our link model, the head and the small AI glasses antenna cost about 4 times in signal strength, and the torso about 100 times more. Together, that is about 400 times weaker, and up to 1,600 times in harsher positions.

The effect on the link is large. With nothing in the way, the signal from the AI glasses reaches the phone about 400,000 times stronger than the background noise. With the phone in a pocket, it arrives only about 1,000 times stronger. A 500 Mbps video stream needs about that 1,000-to-1 margin to get through cleanly. So, AI glasses only one meter from the phone behave like a link 20 meters away in open air, right at the edge of what high-rate video needs (modeled, with the AI glasses sending 17–18 dBm on an 80 MHz channel).

At that edge, the transmitter's own signal error matters more than at any other point. When the link has plenty of margin, a little extra distortion costs nothing. When the body has removed almost all of the margin, every bit of distortion pulls the link down to a slower speed.



The body also changes the antenna itself. A head, hand or even hair close to the antenna detunes it, so part of the power bounces back into the front-end. Engineers measure this mismatch as Voltage Standing Wave Ratio (VSWR); at 3:1 a quarter of the power is reflected. Many front-ends lose linearity under that mismatch, so the signal gets dirtier exactly when the wearer moves. QuantalRF's architecture holds its linearity up to a 3:1 mismatch, independent of supply voltage, so the signal stays clean on the head, in the hand and on the wrist.

Of course, a designer who insists on winning back everything the body takes only needs to send tens of watts from the antenna and draw a few hundred watts from a battery the size of a stick of gum. But few customers want a space heater on their head. Fortunately, the link does not need to win back everything the body takes. It needs enough extra signal to hold its speed as the wearer moves: usually a few dB, the gap between the speed setting it wants and the next one down. There are three ways to find those few dB, as shown in the following table:

Approach	What it Costs on a Wearable	Result
Push today's front-end harder	From 3.3 V, today's front-ends run out of clean output at about 15.5 dBm. Going further needs a boost converter, and each 3 dB more output roughly doubles the front-end's power and heat.	A few dB, at a high battery and heat cost, with a dirtier signal.
Add a second radio	A second front-end and antenna, double the front-end power and about 60% more Wi-Fi chip transmit power.	Still limited by signal quality: 712 Mbps in our model, below one QuantalRF 2nd Gen radio.
Send a cleaner signal	No extra radio and no boost converter. From the same 3.3 V battery, our 2nd Gen sends about 1.8 times as much clean power as the incumbent, and our 3rd Gen about 4 times. Our antenna then gets 2.2 times more of that power out of the glasses.	The link keeps its speed without drawing more power. A cleaner signal also lets the phone use a faster speed setting at the same signal strength.

Packets Sent Again: The Hidden Battery Cost

When a packet arrives with too many errors, it has to be sent again. Each re-send costs the front-end and the Wi-Fi chip the same energy a second time, uses airtime the next packet needed, and delays the video frame it belongs to.

Wi-Fi speed control usually pushes for the fastest setting and accepts a share of lost packets. With a noisy transmitter in a body-blocked link, that share climbs fast. Our simulation streams 30 video frames a second to a phone in a pocket, with the signal fading as the wearer moves (modeled, set up to favor the incumbent):

Result, Phone in a Pocket	Incumbent Front-end	QuantalRF 2nd Gen	QuantalRF 3rd Gen
Packets sent again, 100–150 Mbps	about 20%	about 2%	under 2%
Packets sent again at AI glasses 480 Mbps target	48%	18%	3%
Battery used each hour on re-sends alone (0.9 Wh)	7.8%	0.5%	—
Delay of the slowest 1 in 100 video frames	0.2 s, a visible freeze	0.015 s	—

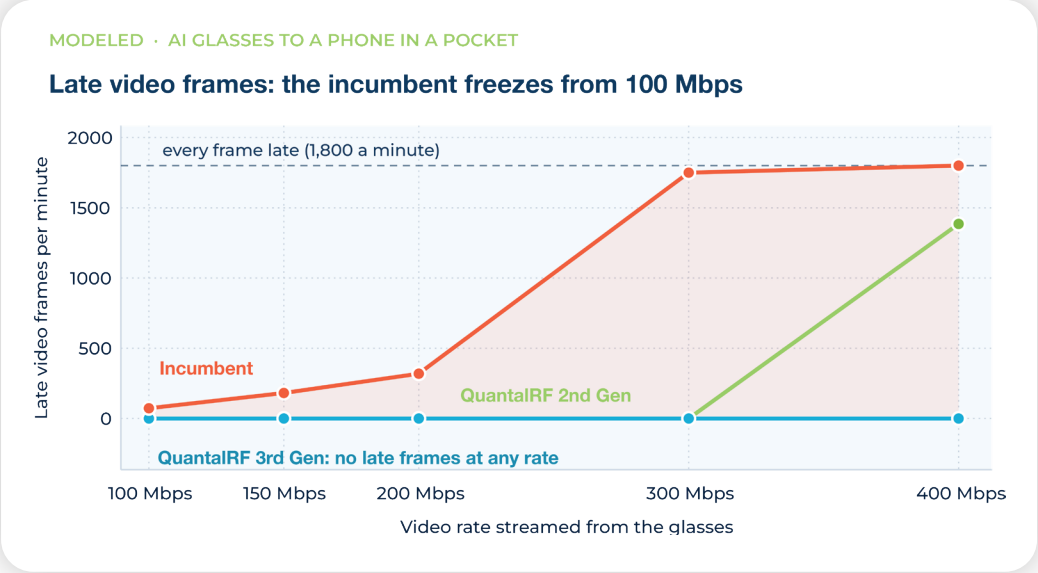


Figure 1. Modeled: AI glasses streaming 30 frames a second on 80 MHz to a phone in a pocket. A frame counts as late when it arrives more than 0.1 s after it should.

Counting re-sends, the front-end and the Wi-Fi chip together, our 2nd Gen delivers each bit for 1.3 to 1.7 times less energy than incumbent front-ends, and our 3rd Gen for 1.8 to 2.4 times less energy. The incumbent's picture starts to freeze at 100 Mbps; ours stays smooth to 300 Mbps with the 2nd Gen, and to 400 Mbps with the 3rd Gen.

Energy Per Bit and Wide Channels

The number that sets battery life is not how much power the front-end draws, but how much energy it spends to move each bit. A front-end that draws the same power on a wider channel moves more bits for the same energy.

Measured: QuantalRF's power draw is the same within 1% across 20, 80 and 160 MHz channels, and within ~3% up to 320 MHz. At 17 dBm in mid-power mode the part draws ~540 mW from 3.3 V at every bandwidth. Doubling the channel doubles the data and halves the energy per bit:

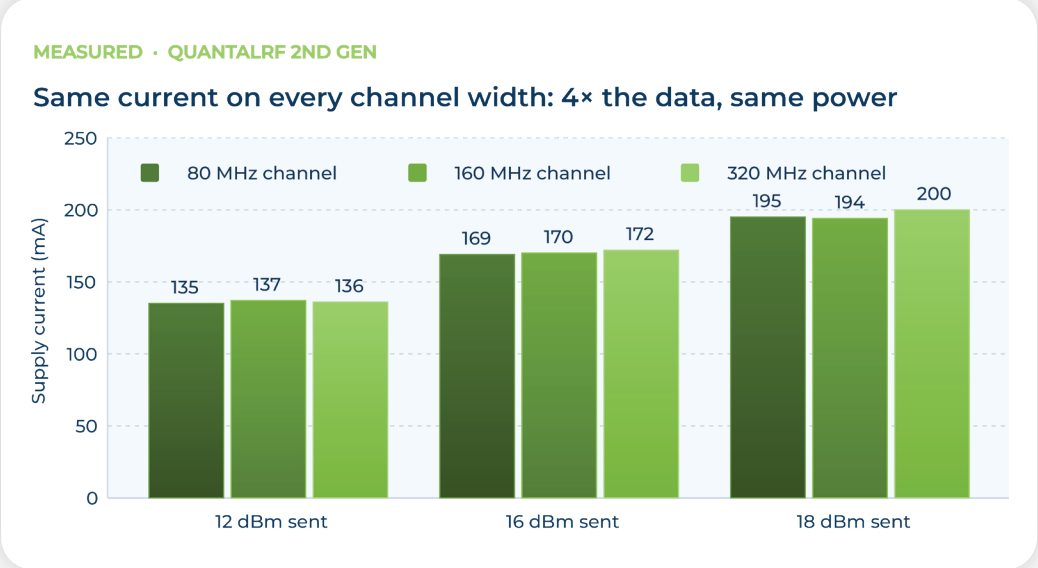


Figure 2. Measured, QuantalRF 2nd Gen, high-power mode, 3.8 V. The 3.3 V mid-power mode of the edge-AI part shows the same flat behavior, within about 3% from 80 to 320 MHz.

Channel width	Top single-stream rate at MCS9	Front-end energy to send 1 GB at 17 dBm
80 MHz	about 480 Mbps	about 9 J
160 MHz	about 960 Mbps	about 4.5 J
320 MHz	about 1,920 Mbps	about 2.3 J

(Peak rates; front-end only. Real throughput is lower, but the ratio between widths holds.)

This improvement in energy only works if the signal stays clean on the wide channel. Today's front-ends keep their clean output on wide channels only in high-power mode, the most power-hungry one. In their low- and mid-power modes, the best published clean output is about 8–9 dBm at 160 MHz and 1.5–7.5 dBm at 320 MHz. Our mid-power mode holds 17 dBm at 0.8% signal error on 160 MHz, matching its 80 MHz performance within 0.3 dB on every channel (measured). A wearable can stream at 80 MHz and switch to 160 MHz, the widest 5 GHz channel many flagship phones support, to upload in half the time, without changing mode or power.

The Radio Must Carry Its Own Linearity

Most Wi-Fi front-ends rely on a digital correction run in the Wi-Fi chip, called digital pre-distortion (DPD): the chip learns how the amplifier bends the signal and pre-bends it the other way. Used well, it buys a few extra dB of margin, but it should not decide whether the product works. The correction has to be trained, it goes out of date when the channel, power, temperature or traffic pattern changes, and, as the next section shows, nobody can guarantee that its training settles on the right answer. Wi-Fi 7 and Wi-Fi 8 make traffic busier, which makes out-of-date corrections more common.

So, the front-end amplifier must carry its own linearity. A front-end that is clean before any correction keeps the link working while the correction is out of date, still training or wrong; the correction then adds margin on top instead of holding the system up. Our 2nd Gen is built this way: linearity in the amplifier, with digital correction as extra margin, not a requirement. Our 3rd Gen goes further. It is completely self-contained: no correction loop and no training, just native linearity with margin to spare. There is nothing to go out of date and no training that can fail to settle.

Why a Training Loop Is the Wrong Tool on a Moving Body

The correction loop has a price, and on a wearable, the battery pays all of it:

- **A second receiver that must stay awake.** To see the distortion, the loop looks for three to five times the channel width: up to about 1 GHz of receiver for a 320 MHz channel, powered on while it trains.
- **Airtime and Wi-Fi chip power.** Each training run takes time on air that video needed, and digital work in the Wi-Fi chip.
- **Stored tables per condition.** Each channel, width, power and temperature can need its own set.

On a phone lying on a desk, conditions hold still long enough for this to work. On a body, they never do. Every step, every turn of the head and every hand raised to the frame changes two things at once: the signal reaching the phone, which in our model swings by about 4 dB from one packet to the next, and the antenna match the amplifier drives into. The first erodes the link's margin; the second changes the amplifier itself. A correction trained a moment ago no longer fits the amplifier it is correcting. The loop must either retrain, spending power again, or keep using a stale correction and send a dirtier signal. Both happen exactly when the body has already removed the link's margin.

Retraining is not even guaranteed to help, and the antenna is the reason. Every amplifier runs some distance below its maximum output, a gap called headroom, so the signal's peaks do not clip. A correction loop behaves well while that headroom holds. But the maximum is not fixed. When a hand or the side of the head detunes the antenna, the amplifier's maximum output can drop, and the headroom the system was set up with quietly shrinks. Near that limit, the loop has no correct answer to settle on. It finds a good answer, then keeps chasing peaks the amplifier can no longer produce, and drifts away from the good answer it found.

In our simulation of a generic CMOS amplifier (Figure 3), a loop running with 9.5 dB of headroom settles comfortably under the target. Take away 1.5 dB of that headroom, which a detuned antenna can do, and the same loop at the same drive finds its best answer after 13 steps and then climbs to 12 times the error. Nobody in the field can see which training step was the best one, so nobody can stop it there.

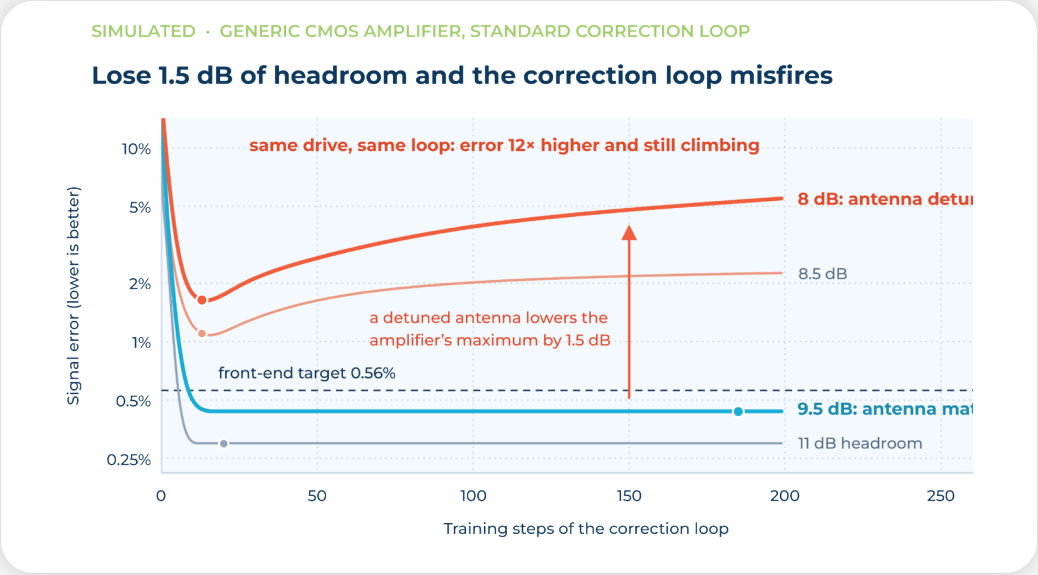


Figure 3. Simulated, from the study first shown at Wi-Fi NOW 2026: a generic CMOS amplifier model with a standard 15-term correction loop, retrained on the same captured packet. Headroom is how far the average output sits below the amplifier's maximum. The 1.5 dB loss to a detuned antenna is an illustration; the real loss depends on the antenna and how it is detuned. These are not QuantalRF measurements.

It gets worse, because no two packets are alike. Each Wi-Fi packet has its own peaks, and the average power says nothing about them. In the same simulation, 40 random packets were sent at exactly the same average power, and each was trained with the same loop (Figure 4). Packets whose peaks stay under the amplifier's maximum settle and stay settled. Packets whose peaks exceed it by more than about half a dB drift away from their best answer, ending with almost four times the error. On a wearable, a detuned antenna can lower that maximum, so more and more packets cross it as the wearer moves, and the loop cannot know which ones.

> The only answer that scales is a **cleaner signal from every milliwatt.**

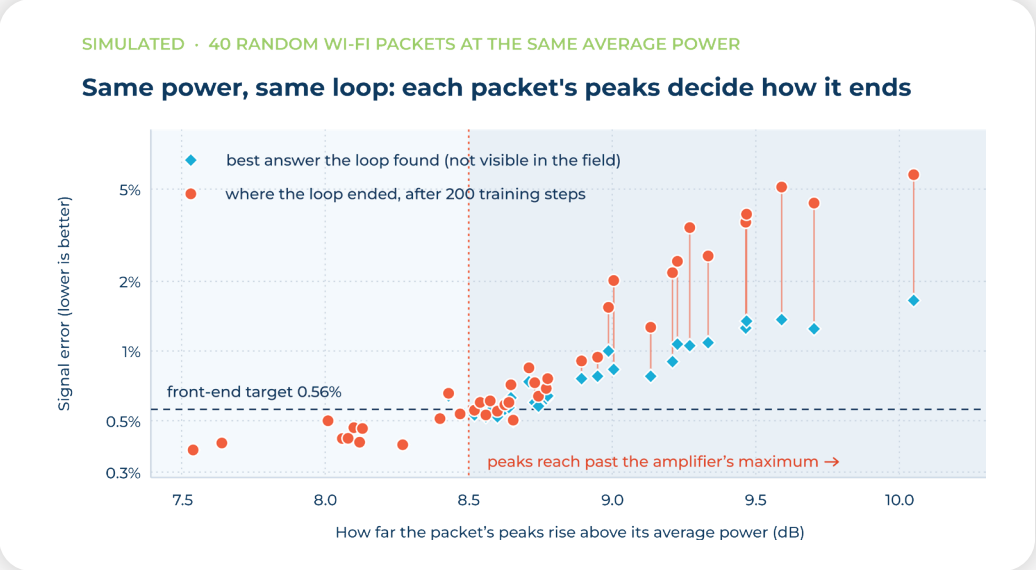


Figure 4. Simulated, same model and loop as Figure 3: 40 random Wi-Fi packets at the same average power, 8.5 dB below the amplifier's maximum, each trained for 200 steps. Not QuantalRF measurements.

What the Wearer Feels When the Loop Misfires

A misfiring correction loop does not show up as an error message. It shows up as a product that disappoints, one small failure feeding the next:

1. A dirtier signal. The correction now adds distortion instead of removing it, just as the body has already taken the link's margin.
2. More packets sent again. The phone rejects more packets, and each re-send spends the same front-end and Wi-Fi chip energy a second time. In our model the incumbent already spends 7.8% of a 0.9 Wh battery every hour on re-sends alone, against 0.5% for our 2nd Gen.
3. More heat against the skin. Every re-send, every retraining run and every milliwatt the misfiring amplifier wastes becomes heat in a frame with no heat spreader, where about 50 mW is enough to warm the skin by 1°C.
4. Losses nobody can account for. None of this appears on a datasheet. The front-end is specified on a test bench with a matched antenna and a freshly trained correction; the cost of a loop that has drifted on a moving head is never measured, so it is never budgeted for.
5. Frozen frames and lag. Late packets hold up the video, and the AI on the phone waits for it. In our model the incumbent already delivers 73 late frames a minute at 100 Mbps.
6. A hot, slow, short-lived product. The wearer does not see a correction loop. They see AI glasses that warm up, lag and need charging before the day is over.

QuantalRF was built to break this chain. A more linear amplifier keeps the signal clean on its own, on the head, in the hand and as the wearer moves. A correction that is late, out of date or wrong costs a little margin rather than the link. Our 3rd Gen, sampling in Q2 2027, needs no correction loop at all.

QuantalRF 2nd Gen with QuantalRF Antennas

QuantalRF's 2nd Gen edge-AI front-end beats today's front-ends on every measure in our wearable model. The following four design choices make the difference on a body-worn device:

Runs directly from 3.3 V. Wearables power their radios from a 3.3 V rail regulated from a single cell. Today's front-ends are built for phones at about 3.8 V. From 3.3 V, they either lose clean output (to about 15.5 dBm in

our model) or need a boost converter, which wastes about 12% of the power. Ours sends 17–18 dBm cleanly straight from 3.3 V, drawing 507 mW at 17 dBm in low-power mode (measured), against 540–650 mW at 3.8 V published for today's best phone front-ends.

26 dB of gain in low-power mode. Gain is how much the front-end amplifies the signal the Wi-Fi chip hands it. Ours amplifies about 400 times in low-power mode, against about 125 times (21 dB) for today's incumbent parts. The Wi-Fi chip therefore needs to supply only about a third of the drive signal, which saves about 50 mW in the Wi-Fi chip's transmit chain while sending (300 mW against 350 mW in our model). That saving is on top of the front-end's own.

Stays clean when the antenna is detuned. Linearity holds up to a 3:1 antenna mismatch, so a hand, hair or the side of the head near the antenna barely affects signal quality.

One small die, off when idle. One chip under 2 mm^2, in a 2 × 2 mm package or delivered as a bare die for module integration, drawing zero current between packets.

The Antenna

QuantalRF's wearable antenna is 25 × 5 × 0.7 mm, fits in the arm of a pair of AI glasses and is made on a standard circuit board or can be printed. It radiates about 90% of the power fed to it, with +1 dBi gain, against 60–70% and about −2.5 dBi for a typical AI glasses antenna. That is about 2.2 times more radiated signal from the same front-end. With our antenna, our 2nd Gen already matches what our 3rd Gen silicon does alone: about 850 Mbps from a pocket, and full-HD video at 35 m.

The Scorecard

AI glasses streaming to a phone in a pocket, 0.9 Wh battery, run the way device makers rate battery life (modeled, set up to favor the incumbent):

Test	Incumbent	QuantalRF 2nd Gen	QuantalRF 3rd Gen	3rd Gen with Our Antenna
Fastest video from one front-end and one antenna	408 Mbps	740 Mbps	850 Mbps	1,034 Mbps
Packets sent again at 480 Mbps	48%	18%	3%	under 1%
Late video frames per minute, 100 Mbps	73	0	0	0
Farthest distance for full-HD video	17.5 m	27 m	35 m	47 m
Battery used by a 1-hour 720p meeting at 20 m	87%	45%	39%	35%
Battery used by a 1-hour 1080p meeting at 20 m	cannot hold it	62%	50%	42%

Battery Life Extended

Based on a 9-hour day with ~30 minutes of video.



Less Heat Against the Face

Nearly every milliwatt a wearable uses ends up as heat in a frame that rests on the skin. Safety standards allow skin contact up to about 43°C for all-day wear (IEC 62368-1 limits), and makers aim lower for comfort, around 39–42°C. AI glasses with room for a heat spreader warm by roughly 2 to 5°C for each extra watt (Matsuhashi et al., 2020; Goodway Technology). The newest, slimmest frames have no room for one, and there the heat stays close to the chip that makes it. Device makers estimate about 1°C for every 50 mW near the component, roughly four to ten times worse than a frame with a spreader.

In our streaming case, AI glasses draw about 210 mW less with our 2nd Gen than with the incumbent, and about 310 mW less with our 3rd Gen. Near the edge of the incumbent's range, the radio alone draws 345 mW less with our 2nd Gen. In a frame without a heat spreader, that is roughly 4 to 7°C cooler against the skin (estimate). That is often the difference between a product that stays under the comfort limit through a long recording and one that has to stop and cool down.

3rd Gen: One Radio Outperforms Two

The usual way to send more data is a second radio: two front-ends, two antennas and two transmit chains in the Wi-Fi chip, each sending its own stream. On a wearable, that doubles the front-end power, adds about 60% to the Wi-Fi chip's transmit power, and needs room for a second antenna and cooling for a second front-end.

It also does not solve the body problem. In our model, an incumbent design with two front-ends and two antennas reaches 712 Mbps from AI glasses to a phone in a pocket. Its limit is signal quality, not energy, so a bigger battery would not raise it.

QuantalRF's 3rd Gen with our antenna reaches 1,034 Mbps from a single front-end and a single antenna in the same case (*modeled*), with no second radio and none of its extra power:

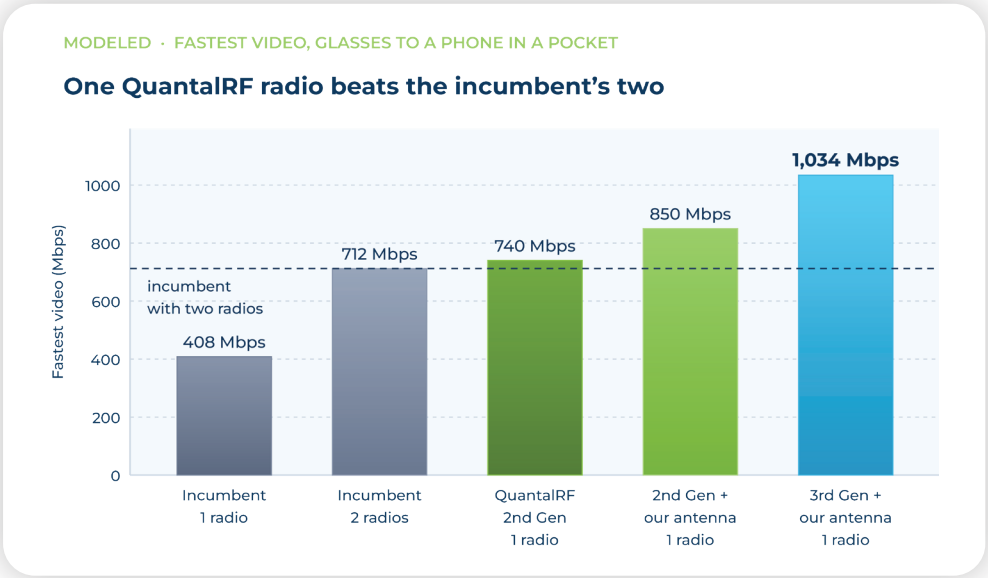


Figure 5. Modeled: glasses to a phone in a pocket, each design on the widest channel it can use cleanly (incumbent 160 MHz, QuantalRF 320 MHz).

Three things combine to get there:

- Twice the amplifier efficiency of our 2nd Gen: half the front-end power for the same output.
- 3.5 dB more clean output, about 2.2 times the power, which raises the clean-power lead over today's parts to 14–50 times.
- Our antenna, worth another 2.2 times in radiated signal on the AI glasses side.

The top rates use the widest Wi-Fi 7 channel, so they need a phone that supports it. Many flagship phones stop at 160 MHz. On that channel, and on the 80 MHz streaming channel for AI glasses, the ranking is the same, but every figure is lower.

Conclusion: Linearity First

The industry's answer to every radio problem has been more: more power, more correction, more radios. On a device worn on the body, each one adds heat and drains a battery that has none to spare. The front-end that wins edge AI is the one that sends the cleanest signal from every milliwatt, on the widest channel, without a loop to retrain.

Four questions to ask of any front-end before it goes into a wearable:

1. What clean output does it reach from 3.3 V, without a boost converter?
2. Does its power draw stay flat as the channel widens to 160 MHz and beyond?
3. Does it stay clean on its own, without relying on a trained correction, including through a 3:1 antenna mismatch?
4. How much gain does it give the Wi-Fi chip, and how much transmit power does that save?

QuantalRF's 2nd Gen part answers all four today, and it is sampling now with Tier-1 device makers. Our 3rd Gen, with twice the amplifier efficiency and no correction loop, delivers more from one radio than today's designs reach with two. Our 3rd Gen part tapes out in Q4 2026, with customer samples available in Q2 2027.

To request samples, reach out to sales@QuantalRF.com.



QuantalRF (USA)
895 Dove Street, Suite 300
Newport Beach, CA 92660
USA
+1 (858) 401-6444

QuantalRF (AG)
Gartenstrasse 10
8002 Zürich
Switzerland
+41 43 5387144

Contact us:
sales@QuantalRF.com

Learn more:
www.QuantalRF.com