PER-RECORDING FIDELITY GUARANTEES FOR WEARABLE ECG COMPRESSION BY ENCODER-SIDE VERIFICATION

Anonymous authors
Paper under double-blind review

ABSTRACT

In a wearable electrocardiograph (ECG), sending one bit over a short-range radio consumes about two thousand times the energy of one arithmetic operation, so battery life depends mainly on compression. Existing compressors are tuned once on a development set and applied to each recording the device later encounters, so fidelity can only be reported as an average, which conceals the recordings that were damaged. On the MIT-BIH ECG Compression Test Database the transform family achieves 23.15x while an automatic reader detects only 82.0% of the QRS complexes. We observe that the encoder retains the original signal, so it can decompress its own output before transmission, compare the beats detected in that reconstruction with the beats it started from, and reject any attempt that moves or loses a beat, retrying at a finer setting. Each recording is then compressed as much as it can tolerate, instead of as much as the least tolerant recording in the database allows, so the coder also achieves higher ratios. Over 162 recordings it obtains 14.81x with all beats preserved, compared with 6.49x at 99.3% beat retention for the published family, and it extends a 225 mA h cell from 304 to 631 days, with verification taking 9 percent of the power budget. Two further measurements limit what any coder can gain. Noise accounts for 87% to 96% of the bits used by a strong neural coder, and that coder is three orders of magnitude less energy-efficient than a linear predictor. Verifying fidelity inside the encoder allows ambulatory monitoring to be audited one recording at a time.

1 INTRODUCTION

A wearable electrocardiograph (ECG) consumes most of its energy in transmission, so reducing the volume of data is the single change that extends battery life by a large factor. Delivering one bit over a short-range radio costs on the order of 100 nJ with protocol and wake-ups included, while one arithmetic operation on a small processor costs about 50 pJ, i.e., about two thousand operations per transmitted bit. This exchange rate frames the problem for any coder, learned or classical: arithmetic is nearly free, and transmitted bits are not.

Compression saves those bits at the expense of fidelity, and the loss is unevenly distributed. When we reimplemented the transform family from which most published compression ratios derive (Tiwari & Falk, 2019) and ran it over the MIT-BIH ECG Compression Test Database (CTDB) (Goldberger et al., 2000), an automatic detector still found 96.9 percent of the QRS complexes at 12.07x. At 23.15x it found 82.0 percent. At 47.42x it found 61.3 percent. The upper part of this range is where the transform family reports its ratios, alongside an error averaged over the recording.

The deeper problem is the evaluation protocol. A setting is chosen on a development set, applied to each recording the device later encounters, and reported as an average over test data (Tiwari & Falk, 2019; Jha & Kolekar, 2022). The average describes the population, while a clinician, or a downstream model, faces one recording at a time, and such a summary cannot identify the recordings that were damaged.

The encoder can prevent this because it still has the original signal. Before transmitting, it can compress the recording, decompress the result, run a detector over the reconstruction, and compare the complexes detected there with those in the buffered signal. When a beat has moved or disappeared,

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

the encoder rejects that attempt and retries at a finer setting, so no data is transmitted until all beats match. Unlike quality-controlled coders, which iterate toward a scalar distortion target (Chen & Itoh, 1998; Miaou & Lin, 2002), the encoder does not assume that a setting chosen in advance generalizes to the recording at hand; it tests the property it must preserve on its own output. The decoder is unchanged and receives only the index of the accepted setting.

Per-recording verification also improves compression, because recordings differ in the distortion they can tolerate: the noise between complexes ranges from 9 to 375 µV across this database (Figure 1), a fixed setting must satisfy its least tolerant recording, and our encoder selected 30 distinct settings across the 162 recordings (Figure 2). The same economics also apply to learned compression (Yildirim et al., 2018; Liu et al., 2024): most of what a strong neural coder encodes on this signal is irreducible noise, and its arithmetic costs orders of magnitude more than the bits it saves. We formulate four research questions; Section 5 presents the corresponding measurements and Section 6 interprets them.

RQ1. How much compression is available when all complexes must be preserved, and how does per-recording verification compare with a single fixed setting?

RQ2. Is the saving retained once the cost of verification is included, measured as battery life on a patch?

RQ3. What fraction of the bit rate is due to noise, and how much room is left for a better model?

RQ4. Can a learned coder justify its computational cost on a wearable?

We contribute a coder that achieves 14.81x with every complex preserved on all 162 recordings and extends a 225 mA h cell from 304 to 631 days; a per-recording noise measurement showing that noise accounts for 87 to 96 percent of the bits a strong convolutional coder uses, which bounds the rate achievable by any model; an energy analysis placing that network three orders of magnitude from breaking even on a patch; and four prior families reimplemented under one set of measures, since published figures use different databases and measures (Tiwari & Falk, 2019) and are not directly comparable (Section 2).

2 Prior work

Work on electrocardiogram compression divides into lossless coders, which reconstruct the recording they were given, and lossy coders, which reconstruct an approximation and report the fidelity achieved. The two literatures report different quantities, so we review them in turn and then note what both fail to measure.

A review of 31 lossless methods (Tiwari & Falk, 2019) identifies linear prediction followed by entropy coding as the most common construction, introduced by Koski (1997), who fits a predictor to each record and encodes the residual. Later work directed at battery-powered devices kept the predictor and simplified the coder (Chen & Wang, 2013; Chen et al., 2013) or combined it with QRS detection so that one pass over the signal serves both (Deepu & Lian, 2015), and the construction remains in use in current real-time coders (Notarianni et al., 2026). Our reimplemented linear predictor follows the same construction, and Section 4 describes the choices it makes.

Transform coders achieve higher ratios. Arnavut (2007) applies the Burrows-Wheeler transform to the residual of a linear predictor and reports 4.3x on the MIT-BIH Arrhythmia Database (MITDB), the highest ratio among the review's 31 methods on that database (Tiwari & Falk, 2019). Dakua & Sahambi (2009) report comparable ratios from a move-to-front transform. Both are offline methods for a structural reason: a block-sorting transform cannot start until the entire block is available, and a streaming device does not accumulate one. Among the methods the same review classifies as capable of real-time operation, the highest reported ratio is 2.84x on the Arrhythmia Database, from a fuzzy predictor with a Huffman region encoder (Chen et al., 2015). The relevant reference point for a real-time coder is therefore 2.84x, not the 4.3x of the offline transform.

Most of the large lossy compression ratios reported for the electrocardiogram come from one construction. A coder transforms the signal, discards the small coefficients, quantizes what remains, and entropy codes it. Discarding small coefficients removes high-frequency content in bulk, in a way no uniform quantizer reproduces, and the reported ratios follow from this. We implement this

family with the reversible integer lifting pair (Sweldens, 1996; Calderbank et al., 1998), which gives a lossless coder when no coefficient is discarded, and Section 5 reports it under the same measures as everything else.

Newer branches report the same averaged measures. Compressed sensing moves cost out of the encoder by sampling below the Nyquist rate (Picariello et al., 2021), with reconstruction quality that worsens as heart rate increases (de Paula Veiga et al., 2026); low-power lossy coders aimed at real-time monitoring quote percentage root mean square difference (PRD) averaged over a database (Fathi et al., 2022; Devindi et al., 2024); deep autoencoders compress the signal end to end (Yildirim et al., 2018; Liu et al., 2024; Bekiryazıcı et al., 2025), with fidelity again reported as an averaged error even where a target error is fixed and the rate varies per record (Bekiryazıcı et al., 2024); and classifiers trained to consume compressed measurements directly (Lal et al., 2026).

Closed-loop control of reconstruction quality also has a history. Chen & Itoh (1998) iterate the quantizer until a chosen PRD is met, Miaou & Lin (2002) deliver a requested quality per signal, and later work replaced PRD with diagnostic distortion measures that weight clinically relevant features (Zigel et al., 2000; Manikandan & Dandapat, 2007). In machine learning, task-aware codecs train the encoder against a downstream model and report the average loss in task accuracy (Ge et al., 2024). All of these control a proxy: a scalar distortion, a weighted feature distance, or a training objective, and Section 5 reports a matched pair in which equal PRD conceals a threefold difference in lost complexes. The coder here instead runs the downstream detector itself and accepts a setting only when the detector output matches beat for beat, with lossless coding as the fallback. None of the families above states which complexes are preserved on a given recording; that is the quantity this coder controls.

Fidelity in this literature is usually reported as the PRD between the original and its reconstruction (Tiwari & Falk, 2019; Jha & Kolekar, 2022). Because it averages over an entire recording, the measure constrains no individual sample, and it is ambiguous: two normalizing conventions are in circulation, and for a recording with a baseline offset they differ substantially. We report the measure under both conventions.

Of the 31 surveyed methods, two state a rate in bits per sample and one evaluates on the CTDB (Tiwari & Falk, 2019), the database assembled to be difficult and the source of our principal result; we therefore report bits per sample and both ratio conventions throughout. What determines battery life on a patch is the balance between the energy saved by transmitting fewer bits and the energy spent computing them, and we measure this balance on the same recordings instead of assuming it.

3 METHOD

The proposed coder has three stages. The first estimates the distortion a recording can tolerate, the second distributes this allowance across frequency bands, and the third confirms on the reconstruction that the allowance was not exceeded. Only the third is unconventional, and the first two define its search space. Throughout, $x_j[t]$ denotes sample t of channel j as the integer produced by the converter, and coding operates on these integers directly, so all reported rates apply to the stored file.

The allowance is estimated from the recording itself. Between the end of one T wave and the start of the next P wave the trace contains no cardiac signal. What remains is the noise particular to the recording, caused by muscle activity, electrode movement, mains interference and the amplifier. From the detected beat positions we extract the window from 0.45 s after one R peak to 0.22 s before the next, subtract the mean of each interval, and denote by σ the standard deviation of the remainder, averaged over channels. By design σ is one value for the recording, not one per channel: the channels are simultaneous views of one heart, and coding the noisier one more coarsely would produce a visible discrepancy between them. Across the database σ varies by a factor of 42, from 9 to 375 µV (Figure 1), so no single allowance can serve the whole database: one chosen for the noisiest recording wastes bits on all the others, while one chosen for the quietest damages the rest.

This allowance is then distributed across bands. We decompose each channel by $L = 5$ levels of the reversible 5/3 lifting pair (Sweldens, 1996; Calderbank et al., 1998), which gives bands $b_0, \ldots, b_L$ from the coarsest approximation to the finest detail. The transform uses only additions and shifts and maps integers to integers, so it reconstructs the recording without error when all step sizes equal one; we verified exact inversion at all signal lengths and decomposition depths. Band k is quantized

with step

$$\Delta_k = \min\big(\max(\lfloor m\,\sigma\,\rho^{\,k}\rceil,\,1),\;c\big), \tag{1}$$

where m scales the estimated noise, $\rho \geq 1$ makes each finer band coarser than the preceding one, and c limits the result. The quantized coefficients are encoded with an adaptive two-sided geometric model, the same entropy coder applied to the linear predictor residual in Section 4, so any difference between the two methods comes from the transform and not from the entropy coding.

Without the limit c, a recording with large σ receives a step size of several hundred microvolts, and quantizing a noisy trace this coarsely produces a staircase that the detector reports as complexes. In this configuration the coder introduced 5.71 false beats per minute, compared with 0.89 for a fixed allowance at a similar rate. The limit therefore allows the coder to adapt where a fixed allowance would waste rate, and prevents adaptation where it would introduce artifacts.

The third stage provides the guarantee. Let $\mathcal{D}$ be a beat detector, and let $\mathcal{S} = (s_1, \ldots, s_N)$ be a list of settings $s = (m, \rho, c)$ ordered from the mildest to the most aggressive. Let $\hat{x}(s)$ be the reconstruction produced under setting s. The encoder accepts s when

$$\mathcal{D}(\hat{x}(s)) = \mathcal{D}(x) \tag{2}$$

holds complex for complex, the two sets being equal in cardinality with each detected position within 50 ms of the corresponding one. Algorithm 1 in Appendix B searches $\mathcal{S}$ for the most aggressive accepted setting, transmits the associated reconstruction, and sends the index of s alongside the coded coefficients.

Two search policies fit this framework, and both exchange rate for computation without affecting fidelity. A linear scan descends $\mathcal{S}$ from the aggressive end and halts at the first accepted setting, which locates the optimal entry in the list but places no limit on the number of evaluations required. Bisection assumes acceptance to be monotone along $\mathcal{S}$ and terminates in $\lceil \log_2 N \rceil$ evaluations, eight for the list used here, at the cost of selecting a milder setting wherever monotonicity is violated. Section 5 evaluates both.

Both policies require $\mathcal{S}$ to be ordered by aggressiveness, and the natural key, the product $m\rho c$, does not provide such an ordering, since raising c while lowering m can leave the product unchanged while the resulting rates differ widely. The list is therefore ordered by the coded size each setting produces, measured once offline on eight recordings and then fixed. The choice is immaterial: bisection under the product ordering also achieves 10.47x. Both policies fall back to entry s_1, which quantizes all bands with step size one and therefore cannot alter a sample or fail Equation 2. This fallback ensures that both policies stop at a setting that preserves the complete beat set, and the guarantee applies to each recording individually, not only on average across a database.

The cost of verification is the additional evaluations, each consisting of a forward transform, a quantization, an inverse transform and a beat detection, all four requiring only additions and comparisons. Section 5 includes this cost in the energy budget instead of excluding it.

4 Experimental setup

Three databases support different parts of the evaluation (Table 3 in Appendix A). The principal result comes from the CTDB (Goldberger et al., 2000), which contains 168 strips selected to be difficult to compress, each with two channels of 5120 samples. Each recording is identified by its source strip and the segment within it, so segments 01 and 02 originate from the same 20 seconds of the same subject. We therefore partitioned by strip, placing 19 source strips on each side with none in both, since alternating segments would leak a calibration from one segment to the other. The method also requires an interval between complexes in which to estimate noise; such an interval was found in 162 of the 168 recordings, and Section 7 discusses the six failures. All reported rates come from the same 162 recordings.

The lossless comparison runs on the MITDB (Moody & Mark, 2001; Goldberger et al., 2000), which is the database the reviewed methods report on (Tiwari & Falk, 2019) and therefore the only one on which our figures can be compared directly with published ones. We pool both directions of the inter-patient split of de Chazal et al. (2004), so each record is encoded by a coder that was not trained on it. The split covers 44 of the 48 records; records 102, 104, 107 and 217 are paced or lack the MLII lead.

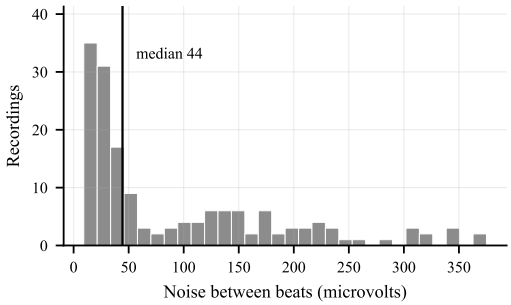

Figure 1: Noise between beats across the 162 recordings varies by a factor of 42, from 9 to 375 µV, so no single error allowance suits the whole range.

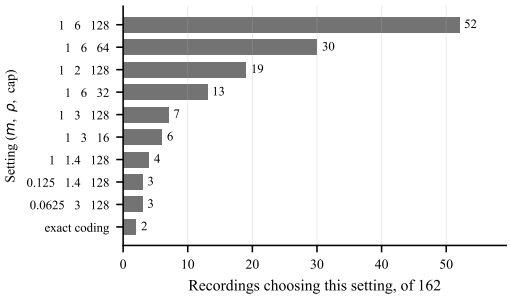

Figure 2: The ten most common of the 30 settings the encoder selected; a fixed-setting method applies one entry of this distribution to all recordings.

The third database, PTB-XL (Wagner et al., 2020), provides the two measurements that limit what any coder can achieve. It contains 21,799 twelve-lead records at 100 Hz in whole microvolts, with folds that keep each patient on one side. A ten-second record at 100 Hz often has too few samples between beats to fit the predictor the noise measurement requires, so we measure the noise floor on 371 of the first 800 records of its test fold. The same measurement on MITDB runs on 21 of the 22 records held out in one direction of the split. PTB-XL also provides the one test of whether a per-recording allowance protects a downstream reader, a five-label diagnostic model trained on 6,000 records and evaluated on 250 held-out records, each compressed before scoring.

We rebuilt four families of baselines and ran them on the same recordings, because the reviewed methods report on different databases and different subsets of them (Tiwari & Falk, 2019), so their published ratios are not directly comparable to ours. Differencing predicts the previous sample and second differencing predicts along the line through the last two, both coding the residual with a Golomb-Rice code whose parameter is chosen per block of 128 samples and transmitted, following FLAC and Shorten. Because a Rice code does not attain the entropy, we count the bits the code actually produces, not the entropy of the residual. Linear prediction fits coefficients to the record being encoded and quantizes them to 12 fractional bits, forming the prediction in integer arithmetic so that the decoder reproduces it exactly; the coefficients are transmitted and included in the rate. Prediction orders from 1 to 32 were tested and the best is reported. A multivariate variant predicts each channel from the history of all channels and from those already encoded at the same instant, with coefficients estimated once on the training partition and distributed with the coder, so it incurs no per-record header. The fourth family is the transform construction of Section 3 with one difference: its step list is fixed in advance instead of being derived from the recording being encoded. Verification is thus the only difference between it and the proposed coder.

Ratios are computed with respect to both reference widths in circulation. One is the stored file, which for the 212 format is 12 bits per sample because it packs two samples into three bytes. The other is the native sample width, 12 bits for the compression test database and 11 for the Arrhythmia database. Using the incorrect reference changes a ratio by 9 percent, more than the difference between several entries in Table 2. Appendix D verifies with an arithmetic coder that the predicted rates are attainable, to within 0.0001 bits per sample.

We report fidelity in three ways, the first of which is not conventional. A per-sample guarantee is a statement about the worst-case error, not the mean, so we compute this error over all samples of all recordings and express it as a multiple of the noise estimated for the recording concerned. Beat agreement is the proportion of complexes present in the original that a detector still locates in the reconstruction, reported together with the rate of spurious complexes. It uses the same detector the encoder verifies against, so the guarantee is conditional on that detector. PRD is reported under both of the conventions described in Section 2, so that our figures can be compared under either.

5 Results

Over the 162 recordings for which noise could be estimated, the proposed coder achieved 14.81x at 0.810 bits per sample with no complex lost. Table 1 compares this with three families that fix a single

Table 1: Compression at matched beat preservation, over 162 recordings

Beats retained	Transform, one setting	Flat step	Transform, noise budget
99.5 percent	not reached	6.62x	7.78x
99.0 percent	6.49x	7.03x	7.78x
98.0 percent	8.78x	15.98x	11.76x
95.0 percent	12.07x	15.98x	17.66x
90.0 percent	16.55x	15.98x	25.83x
All beats	**14.81x** full scan, **10.47x** bounded search		

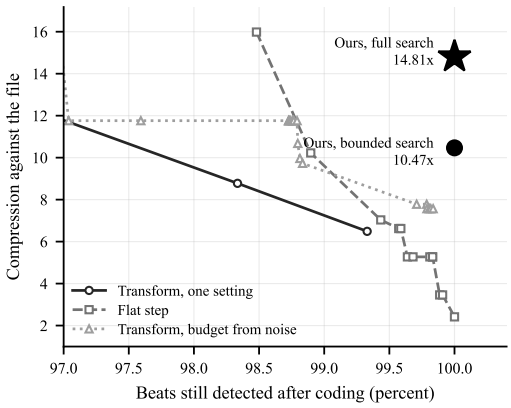

Figure 3: Compression at matched beat preservation. Fixed settings lose compression as the required fidelity increases; verification achieves 14.81x with no beats lost, or 10.47x within eight passes. The axis starts at 97 percent; Table 1 covers lower values.

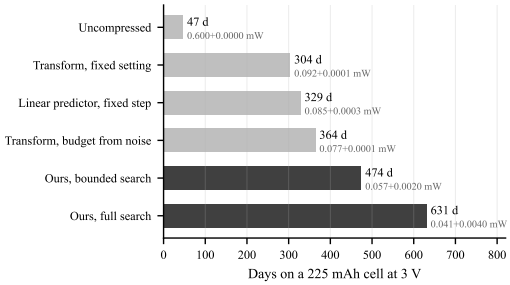

Figure 4: Days of operation from a 225 mA h cell. Verification consumes 0.0040 mW, 9 percent of device power, and extends life from 304 to 631 days.

setting for the whole database, evaluated at five levels of beat preservation, and Figure 3 presents the same comparison as a frontier. The ranking depends on the required fidelity. When 99.5 percent of complexes must be retained, no fixed setting exceeded 7.78x, and the transform family from which most published ratios derive (Tiwari & Falk, 2019) failed to attain this level at any setting. Our coder retained the complete beat set at 1.90 times the compression of the best fixed setting and 2.28 times that of the transform family evaluated at 99.0 percent. At the permissive end the ranking reverses: with 98.0 percent of complexes retained, a flat quantizer achieved 15.98x, more than the 14.81x of the proposed coder.

The additional compression comes from adaptation, not from an improved transform. The noise between complexes varied by a factor of 42 across these 162 strips (Figure 1), so a coder fixed in advance must encode the whole database at the rate its most sensitive member requires. Per-recording measurement regains this margin. The encoder selected 30 distinct settings (Figure 2), with a median worst-case error of 4.24 times the noise already present and a median PRD of 14.00.

Choosing among these settings requires evaluations. Scanning the list from the aggressive end until a setting is accepted required 16.4 evaluations on average, with a long tail (the 95th percentile is 76 evaluations). Bisection over the same list needs at most eight evaluations and produced 10.47x in place of 14.81x. No beats were lost by either variant on any recording.

A matched pair illustrates why we report retained complexes instead of the percentage error the lossy literature quotes. At 5.72 percent root mean square difference the transform family retained 98.3 percent of complexes; at 5.80 percent, only 0.08 points higher, a flat quantizer retained 99.4 percent. Two methods that the measure treats as equivalent therefore differ threefold in complexes lost, because a transform coder removes high-frequency content across the band, taking the edges of the QRS complex with it, whereas a quantizer distributes error uniformly and preserves those edges.

Compression is useful on a patch only if it extends operating life, so we converted each method into days of operation from a 225 mA h cell at 3 V (Figure 4). Uncompressed transmission depleted the cell in 47 days, the transform family with a fixed setting lasted 304 days, and the proposed coder lasted 631 days, falling to 474 days when the search was limited to eight evaluations. Verification is affordable because none of its four operations requires a multiplication. At 16.4 evaluations the arithmetic consumed 0.0040 mW compared with 0.041 mW for the radio, so verification accounted for 9 percent of device power while enabling the reduction of the remaining 91 percent. Neither hardware constant was measured on a board, so both were varied over a wide range: changing radio energy from 15 to 500 nJ per bit and processor energy from 2 to 200 pJ per operation left the ordering of the five methods unchanged.

These savings are limited by a noise floor. A coder that reconstructs a recording exactly must also encode its noise exactly, and noise is incompressible by definition, so the cost of the noise in a given recording limits the saving available to any model. For each recording we estimated a linear predictor on the interval between beats and computed the entropy of the unpredicted remainder. Noise accounted for 87.1 percent of the bits consumed by a strong convolutional coder on PTB-XL and 96.0 percent on MITDB (Figure 5). The correlation across recordings between coded size and estimated noise cost was 0.828, so the noise level of a recording largely determines its coded size. Only 0.669 bits per sample on PTB-XL and 0.145 bits per sample on MITDB remained unaccounted for.

Under a white-noise model of this remainder, the estimated floor lies above the rate our coder already achieves on all recordings; since no lossless coder can code below the true floor, the whiteness assumption must fail. Electrocardiographic noise is correlated between adjacent instants, and a predictive coder removes part of it as a side effect. Three further assumptions act in the same direction: the intervals between complexes retain a little atrial activity and the end of the T wave, the entropy expression assumes a Gaussian, the maximum-entropy distribution at fixed variance, and the predictor is linear, so structure it misses is treated as incompressible. All three raise the estimated floor, so the estimated margin above it is a lower bound.

Whether a learned model can exploit this margin economically is a separate question, which we address through execution cost. Relative to a linear predictor on the compression test database, the network saved 101 bits per second while requiring 206,906,000 additional multiply-accumulate operations per second. This exchange breaks even only where a transmitted bit costs 2042 nJ for each 1 pJ consumed by an operation, and currently available parts operate near 2. Reducing network size does not close a discrepancy of this magnitude, since even our smallest trained model requires about four orders of magnitude more arithmetic than an eight-tap linear predictor. Where transmission is not the binding constraint, however, the network is better by the margin shown in Table 2. For lossless coding of the MITDB it achieved 3.24x combined over both folds, exceeding the 2.84x that the review attributes to the best method it classifies as capable of real-time operation (Chen et al., 2015; Tiwari & Falk, 2019), although it did not match the 4.3x of the offline Burrows-Wheeler method (Arnavut, 2007). Our reimplemented linear predictor also exceeded the real-time figure, with 3.07x at 12 multiplications per sample.

Four further modifications produced no useful gain: added expert capacity, a wider receptive field, an explicit cross-beat pointer, and reverse water-filling. Appendix C reports each in bits per sample relative to a matched control, together with the mechanism that most likely explains it.

A per-recording allowance also failed to protect a diagnostic classifier. We trained a five-label model on 6,000 PTB-XL records and evaluated it on 250 held-out records, compressing the records before scoring. A fixed allowance of 60 µV resulted in a mean AUROC of 0.858 at 1.024 bits per sample, versus 0.854 at 1.052 bits for an allowance of three times each recording’s own noise; 120 µV produced 0.847 at 0.846 bits, versus 0.808 at 0.890 bits for six times the noise. The third pair showed the opposite ordering, 80 µV at 0.842 versus 0.847 for four times the noise, but in all three cases the noise-proportional allowance used more bits. Lossless coding left the mean AUROC at its uncompressed value of 0.876, so the loss is caused by the allowance and not by the codec.

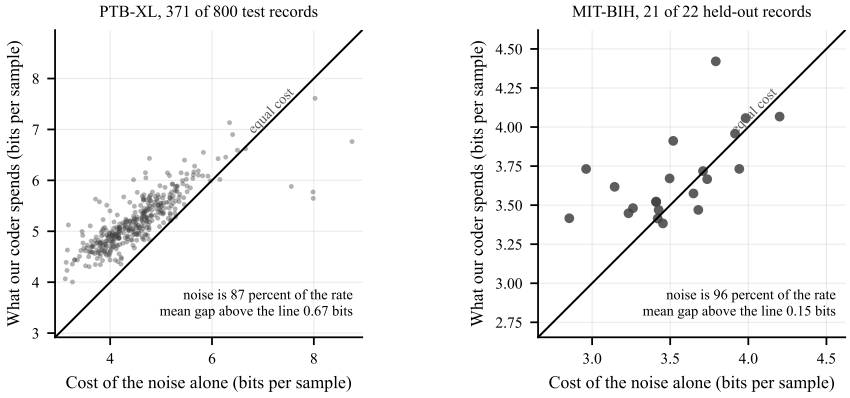

Figure 5: Coded size against estimated noise cost. Each point is one recording; the line marks equal cost. Noise accounts for 87 percent of the rate on PTB-XL and 96 percent on MITDB.

Table 2: Lossless coding of the MIT-BIH Arrhythmia Database

Method	Class	Ratio
Fuzzy predictor, Huffman (Chen et al., 2015)	real time	2.84x
Burrows-Wheeler (Arnavut, 2007)	offline	4.3x
Linear predictor, rebuilt here	real time	3.07x
Our neural coder	real time	3.24x

6 Discussion

We return to the four questions of Section 1. Per-recording verification preserves all detected complexes at 14.81x on all 162 recordings, 1.90 times the best fixed setting under the same requirement, while the transform family behind most published ratios (Tiwari & Falk, 2019) has no setting that meets it (RQ1). The saving is retained after its own cost: verification consumes 9 percent of device power and enables the reduction of the remaining 91, so a cell lasts 631 days in place of 304 (RQ2). Noise accounts for 87 to 96 percent of the bit rate, and the remaining margin of a few tenths of a bit per sample is an upper bound, because each assumption in the estimate raises the estimated noise floor (RQ3). A learned coder cannot cover its arithmetic cost on a patch: breaking even would require a transmitted bit to cost two thousand times as much as one operation, while on current parts the factor is close to two (RQ4).

The first of these results does not come from a better transform. Our transform is the standard reversible lifting pair, our entropy coder is the one the baselines use, and at 98.0 percent of complexes a flat quantizer still achieves a higher ratio than our coder. What changes is that the guarantee applies to the recording instead of the database. A compression ratio is a statistic over a database, but a clinician interprets it as a prediction about the individual recording under examination. The two quantities differ whenever recordings do, so the database-level figure describes the recordings the method was tuned on and few others. Moving the guarantee to the recording costs one additional decode inside the encoder.

The measures in common use help explain why this loss has attracted little attention. PRD averages within a recording, and a reported ratio averages across a database, so damage concentrated in a few recordings is attenuated twice before it is reported. In our matched pair, two settings the measure places within 0.08 points of each other differ threefold in complexes lost, because squared error is insensitive to where in the spectrum an error falls, while a QRS detector is not. Reporting fidelity this way simply cannot show the damage in Table 1, so a per-recording guarantee has to be provided directly, and the proposed encoder does so.

Verifying each recording is affordable only because an evaluation requires additions and comparisons while a transmitted bit requires a radio wake-up, which on current parts costs about two thousand times as much. This is a property of present hardware, not a theorem: if compute and radio energy became comparable, the 9 percent overhead would grow until verification cost more than the bits it saves.

One result is at odds with the principle we advocate. On PTB-XL, at matched rate, an allowance derived from each recording's own noise did not preserve a diagnostic classifier better than a fixed allowance, and in two of three comparisons preserved it less well. Per-recording adaptation is therefore not beneficial in itself. Allocating error in proportion to noise assigns the largest allowance to the noisiest recordings, which are the recordings whose diagnostic features are weakest, so the mechanism that protects a detector can harm a classifier. This observation applies beyond our design: verifying a property does not verify a correlate of that property, and a coder that checks beat positions provides evidence about beat positions only.

Two measurements limit what a better model could gain. If 87 to 96 percent of a coded electrocardiogram is noise that no exact coder may discard, the remaining margin is a few tenths of a bit, and a network cannot capture it on a patch without consuming a thousand times the energy those bits would save. For this signal at this scale, little is left to gain from better modeling and much more from deployment. In our view, the deployment result is the more important of the two.

The numerical bound is specific to electrocardiography; the mechanism is not. Whenever the output of a lossy codec is consumed by an algorithm instead of a human, the encoder can run that algorithm on its own output and reject any setting under which the algorithm fails. The electrocardiogram suits this construction because its consumer is a beat detector requiring only additions; the approach transfers wherever the downstream algorithm is inexpensive relative to the channel.

7 Limitations

The energy figures are operation counts multiplied by nominal constants, not measurements taken from hardware. We compare methods instead of reporting absolute values, so the analysis remains valid as long as the ranking is stable over the realistic range of both constants, which it is. An engineering decision would still require the measurement.

Second, the guarantee itself is relative to a beat detector: a deployed system would verify using the detector of its downstream reader and obtain a correspondingly different guarantee. Beat position is also only one clinical property, and the PTB-XL experiment shows that verifying one property does not protect another; directly verifying a classifier is an extension we have not undertaken. Verification also introduces latency, since the encoder buffers a recording, encodes it up to eight times, and transmits only after acceptance, which suits burst transmission but not sample-by-sample streaming.

A more serious limitation is that the method requires a measurable interval between complexes and does not always find one: six of the 168 strips provide at most one usable interval, and a ten-second PTB-XL record at 100 Hz provides too few noise-only samples on 429 of the first 800 test records, a constraint that binds at high heart rates. A deployed coder would need a fallback, and the obvious one, coding such recordings without loss, uses the most rate on the very records where an arrhythmia is most likely. Finally, the principal result depends on 162 short strips from one database; longer monitoring would change the balance between per-record headers and payload, which we have not measured.

8 Conclusion

Because the encoder retains the original recording, it can compare what it is about to transmit with what it received; this converts a database-level statement into a per-recording one and gives each recording the distortion it tolerates. The family behind the published ratios (Tiwari & Falk, 2019) lost 0.7 percent of complexes at 6.49x; per-recording verification achieved 14.81x with every beat intact and extended cell life from 304 to 631 days. We release the coder, the four baselines, the noise measurement, and the evaluation code, so the comparison can be reproduced and later methods evaluated under the same measures.

AI use statement

In this work, we used generative AI tools for drafting and editing the manuscript text and for literature retrieval while assembling the related-work discussion. We have not used generative AI tools

for research ideation, for the design or execution of the experiments, or for writing the code, and the remaining disclosure categories are not applicable to this work. We have reviewed all AI-assisted text; every reference found this way was verified against its published record, and every number and claim was checked by the authors against the underlying experimental records. We take responsibility for the final content of this work, including text produced with the aid of generative AI.

Ethics statement

This study uses only publicly available, de-identified recordings from PhysioNet (the MIT-BIH databases and PTB-XL); no data were collected from human subjects for this work. The proposed guarantee is defined relative to a beat detector and, as Section 6 reports, does not by itself protect downstream diagnostic classifiers, so clinical deployment would require validation beyond the scope of this paper.

Reproducibility statement

We release the coder, the four reimplemented baselines, the noise measurement, and the evaluation code at `https://anonymous.4open.science/r/verified-ecg-compression-B047/`. Section 4 specifies the databases, the record subsets, the train/test partitions, and the reference widths used for every reported ratio, and all rates are also given in bits per sample so that later methods can be compared without reconstructing our conventions.

References

Ziya Arnavut. ECG signal compression based on Burrows-Wheeler transformation and inversion ranks of linear prediction. *IEEE Transactions on Biomedical Engineering*, 54(3):410–418, 2007.

Tahir Bekiryazıcı, Gürkan Aydemir, and Hakan Gürkan. Electrocardiogram signal compression using deep convolutional autoencoder with constant error and flexible compression rate. *IRBM*, 45:100859, 2024.

Tahir Bekiryazıcı, Mehmet Damkacı, Gürkan Aydemir, and Hakan Gürkan. A novel multichannel sparse convolutional autoencoder for electrocardiogram signal compression. *Journal of Electrocardiology*, 93:154125, 2025.

A. R. Calderbank, Ingrid Daubechies, Wim Sweldens, and Boon-Lock Yeo. Wavelet transforms that map integers to integers. *Applied and Computational Harmonic Analysis*, 5(3):332–369, 1998. doi: 10.1006/acha.1997.0238.

J. Chen and S. Itoh. A wavelet transform-based ECG compression method guaranteeing desired signal quality. *IEEE Transactions on Biomedical Engineering*, 45(12):1414–1419, 1998.

Szu-Lin Chen and Jhih-Guang Wang. VLSI implementation of low-power cost-efficient lossless ECG encoder design for wireless healthcare monitoring application. *Electronics Letters*, 49(2): 91–93, 2013.

Szu-Lin Chen, Guan-An Luo, and Ting-Lan Lin. Efficient fuzzy-controlled and hybrid entropy coding strategy lossless ECG encoder VLSI design for wireless body sensor networks. *Electronics Letters*, 49(17):1058–1060, 2013.

Szu-Lin Chen, Min-Chun Tuan, Ting-Kuang Chi, and Ting-Lan Lin. VLSI architecture of lossless ECG compression design based on fuzzy decision and optimisation method for wearable devices. *Electronics Letters*, 51(18):1409–1411, 2015.

Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley, 2nd edition, 2006. doi: 10.1002/047174882X.

Sarada Prasad Dakua and Jyotinder S. Sahambi. Lossless ECG compression for event recorder based on Burrows-Wheeler transformation and move-to-front coder. *International Journal of Recent Trends in Engineering*, 1(3):120–123, 2009.

Philip de Chazal, Maria O'Dwyer, and Richard B. Reilly. Automatic classification of heartbeats using ECG morphology and heartbeat interval features. *IEEE Transactions on Biomedical Engineering*, 51(7):1196–1206, 2004. doi: 10.1109/TBME.2004.827359.

Thaís de Paula Veiga, Víctor Araújo de Paiva, and Alexandre Rodrigues Farias. Comparative analysis of the quality of ECG signal reconstruction by compressed sensing at different heart rates. *Biomedical Signal Processing and Control*, 122:110426, 2026.

Chacko John Deepu and Yong Lian. A joint QRS detection and data compression scheme for wearable sensors. *IEEE Transactions on Biomedical Engineering*, 62(1):165–175, 2015.

Isuri Devindi, Sashini Liyanage, Titus Jayarathna, Janaka Alawatugoda, and Roshan Ragel. A novel ECG compression algorithm using pulse-width modulation integrated quantization for low-power real-time monitoring. *Scientific Reports*, 14, 2024.

Islam S. Fathi, Mohamed Abd Allah Makhlouf, Elsaeed Osman, and Mohamed Ali Ahmed. An energy-efficient compression algorithm of ECG signals in remote healthcare monitoring systems. *IEEE Access*, 10:39129–39144, 2022.

Xingtong Ge, Jixiang Luo, Xinjie Zhang, Tongda Xu, Guo Lu, Dailan He, et al. Task-aware encoder control for deep video compression. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 26036–26045, 2024.

Ary L. Goldberger, Luis A. N. Amaral, Leon Glass, Jeffrey M. Hausdorff, Plamen Ch. Ivanov, Roger G. Mark, Joseph E. Mietus, George B. Moody, Chung-Kang Peng, and H. Eugene Stanley. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, 2000. doi: 10.1161/01.CIR.101.23.e215.

Chandan Kumar Jha and Maheshkumar H. Kolekar. Electrocardiogram data compression techniques for cardiac healthcare systems: A methodological review. *IRBM*, 43(3):217–228, 2022.

Antti Koski. Lossless ECG encoding. *Computer Methods and Programs in Biomedicine*, 52(1): 23–33, 1997.

Bharat Lal, Vahid Abolghasemi, Raffaele Gravina, Derek O'Keeffe, and Davood Roshan. Compressed learning for real-time ECG signals classification. *IEEE Sensors Journal*, 26:10891–10904, 2026.

Xingyu Liu, Lesong Zheng, Yuhao Cheng, Wenliang Zhu, Jiacheng Li, Jiangxu Liu, Sisi Lu, and Binquan You. End-to-end ECG signal compression based on temporal information and residual compensation. *Circuits, Systems, and Signal Processing*, 44:2779–2801, 2024.

M. Sabarimalai Manikandan and S. Dandapat. Wavelet energy based diagnostic distortion measure for ECG. *Biomedical Signal Processing and Control*, 2(2):80–96, 2007.

Shaou-Gang Miaou and Chih-Lung Lin. A quality-on-demand algorithm for wavelet-based compression of electrocardiogram signals. *IEEE Transactions on Biomedical Engineering*, 49(3): 233–239, 2002.

George B. Moody and Roger G. Mark. The impact of the MIT-BIH arrhythmia database. *IEEE Engineering in Medicine and Biology Magazine*, 20(3):45–50, 2001. doi: 10.1109/51.932724.

Luca Notarianni, Anna Sabatini, Giulia Di Tomaso, and Luca Vollero. Real-time lossless compression for biomedical signal processing. *IEEE Access*, 14:130224–130234, 2026.

Francesco Picariello, Grazia Iadarola, Eulalia Balestrieri, Ioan Tudosa, and Luca De Vito. A novel compressive sampling method for ECG wearable measurement systems. *Measurement*, 167: 108259, 2021.

Wim Sweldens. The lifting scheme: A custom-design construction of biorthogonal wavelets. *Applied and Computational Harmonic Analysis*, 3(2):186–200, 1996. doi: 10.1006/acha.1996.0015.

Abhishek Tiwari and Tiago H. Falk. Lossless electrocardiogram signal compression: A review of existing methods. *Biomedical Signal Processing and Control*, 51:338–346, 2019.

Patrick Wagner, Nils Strodthoff, Ralf-Dieter Bousseljot, Dieter Kreiseler, Fatima I. Lunze, Wojciech Samek, and Tobias Schaeffter. PTB-XL, a large publicly available electrocardiography dataset. *Scientific Data*, 7(1):154, 2020. doi: 10.1038/s41597-020-0495-6.

Ozal Yildirim, Ru San Tan, and U. Rajendra Acharya. An efficient compression of ECG signals using deep convolutional autoencoders. *Cognitive Systems Research*, 52:198–211, 2018.

Y. Zigel, A. Cohen, and A. Katz. The weighted diagnostic distortion (WDD) measure for ECG signal compression. *IEEE Transactions on Biomedical Engineering*, 47(11):1422–1430, 2000.

A Database summary

Table 3 lists the recordings behind each experiment; the subsets and splits are described in Section 4.

Table 3: Recordings used in each experiment; counts refer to records actually evaluated, not to full database size

Database	Records	Rate	Width	Role
CTDB	162 of 168	250 Hz	12 bit	main result
MITDB	44 of 48	360 Hz	11 bit	lossless
MITDB	21 of 22	360 Hz	11 bit	noise floor
PTB-XL	371 of 800	100 Hz	16 bit	noise floor

B Encoding algorithm

Algorithm 1 lists the two search policies of Section 3 in full. SCAN produces the rate we report, and BISECT limits the number of passes to eight.

Algorithm 1 Encoding with verification. SCAN gives the rate we report, and BISECT limits the number of passes to eight

Require: recording x, detector $\mathcal{D}$, settings $\mathcal{S}$ ordered by measured cost, mildest first

$\sigma \leftarrow$ noise in the intervals between the beats of x
$P \leftarrow \mathcal{D}(x)$
function ACCEPTS(s)
    $\hat{x} \leftarrow$ decode(encode(x, s, σ))
    return $\mathcal{D}(\hat{x}) = P$ beat for beat
function SCAN
    for $k \leftarrow N$ **down to** 2 **do**
        if ACCEPTS(s_k) **then return** s_k
    return s_1 ▷ exact coding
function BISECT
    $lo \leftarrow 2, \;\; hi \leftarrow N, \;\; best \leftarrow s_1$
    while $lo \leq hi$ **do**
        $mid \leftarrow \lfloor (lo + hi)/2 \rfloor$
        if ACCEPTS(s_{mid}) **then**
            $best \leftarrow s_{mid}, \;\; lo \leftarrow mid + 1$
        else
            $hi \leftarrow mid - 1$
    return $best$
$best \leftarrow$ SCAN ▷ or BISECT to limit the passes to eight
return encode(x, $best$, σ) and the index of $best$

C Modifications that did not help

Table 4 reports the modifications in bits per sample relative to a matched control. For each we also give the likely mechanism, since a mechanism is more informative for later work than a number by

Table 4: Modifications with little or no benefit, in bits per sample relative to a matched control; positive entries are losses

Change	Control	Change in bits
128 specialists, 18x the weights	4.087	$+0.039$
Load balance, 5.0 to 0.05	4.126	$+0.001$
Context, 4 to 18 beats	3.654	-0.001
Context, 4 to 57 beats	3.654	$+0.109$
Beat pointer, PTB-XL	4.069	-0.058
Beat pointer, MITDB	3.656	-0.031

itself. Additional capacity does not help: a mixture of 128 experts with 34.7 million parameters, 1.9 million of them active at any position, was outperformed by a dense network of equal running cost, and relaxing the load balance from 5.0 to 0.05 changed the rate by only 0.001 bits, so the balancing constraint was not the limiting factor. A wider receptive field does not help either. Extending it from 4 beats to 18 left the rate unchanged, and extending it to 57 cost 0.109 bits. The latter model also finished with the higher training loss, so it is probably undertrained rather than harmed by the wider context.

Cross-beat structure exists, but we could not exploit it. The residual at an instant was 41.1 percent predictable from the residual at the corresponding point of the preceding beat, compared with 0.06 percent for a control offset with no beat alignment, a structure corresponding to 0.366 bits once the beat positions have been transmitted. Supplying the network with this pointer explicitly recovered less than a tenth of it. A predictor estimated on pooled residuals probably captures heteroscedasticity the model already handles through a wider output distribution, so the residual figure likely overestimates what an explicit pointer can recover.

Reverse water-filling also failed, for a reason that indicates where the rate is spent. Rate-distortion theory allocates no rate to a component below the water level (Cover & Thomas, 2006), and setting this level from each band's measured noise is formally the same as the manual allocation in Section 3. It achieved 7.62x at 95 percent of complexes retained, compared with 17.66x for the allocation we adopt.

D Entropy-coder validation

A predicted rate is only meaningful if a coder can attain it, so we encoded a segment of each configuration with an arithmetic coder and compared the bytes written with the bits predicted; the discrepancy was 0.0001 bits per sample. This comparison also revealed an implementation error. Supplying the coder with the minimum and maximum of the record caused it to renormalize over a smaller alphabet than the model assumed, and it wrote 0.031 bits per sample fewer than predicted. Since no coder can outperform its own model, the sign of the discrepancy identified the error.