ECSEQ-0: Label-Free Electrochemical Impedance Sequencing by Synthesis

Proprium Bioscience
M. Vaden, J. Ready, E. Goldstein
https://propriumbioscience.com/research/ecseq-0

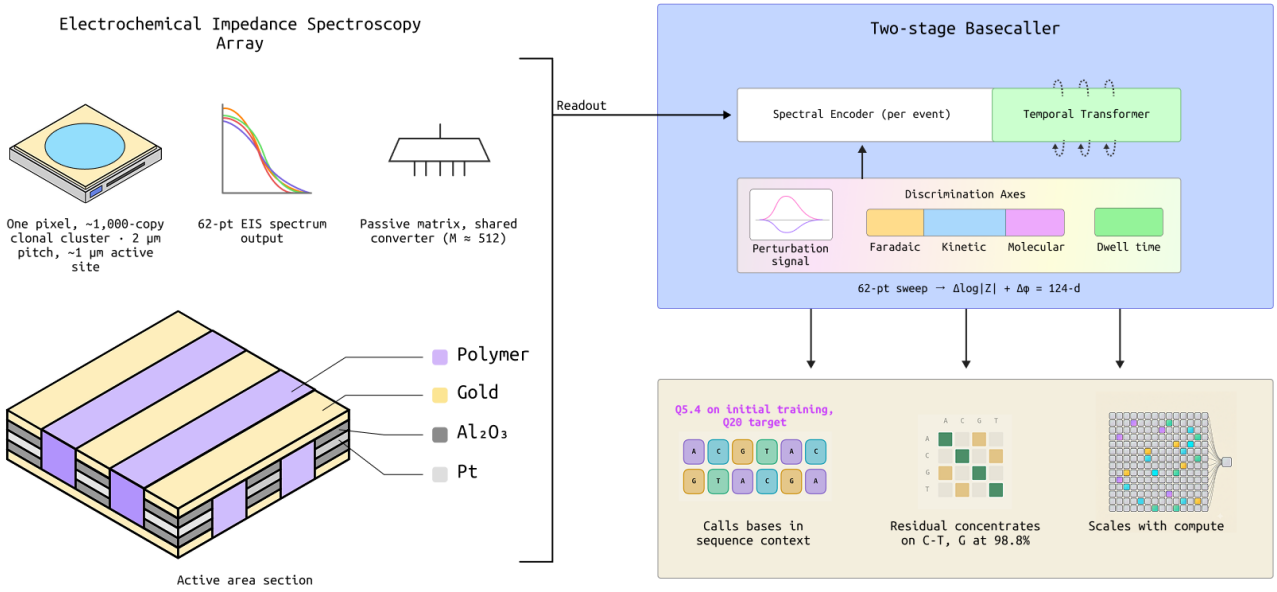

Fig. 1: ECSEQ-0 end to end. A passive electrode array performs localized impedance spectroscopy at each pixel; a two-stage decoder turns the 62-point sweep into a base call, with the discrimination carried by three frequency-distinct axes rather than by a label.

Abstract. A dense array of microelectrodes that measures its own interfacial electrochemistry, read by a model trained to interpret that measurement, is a general-purpose biological instrument. The array is passive and fixed and the discrimination lives in the decoder, so a new assay means retraining rather than refabricating, and the instrument improves as models improve rather than as fabrication improves. We develop such an array against its hardest application, label-free and wash-free DNA sequencing by synthesis. As a polymerase extends a surface-bound strand, the interfacial impedance shifts in a base-dependent way; measured by EIS at each pixel, with no label to image and no reagent exchanged between incorporations, that shift separates into three frequency-distinct axes (kinetic, molecular, faradaic) carrying enough information to discriminate all four bases, which a per-event spectral encoder and a temporal transformer read. On physically simulated data the decoder reaches 71.2% per-base (Q5.4) at L = 30 against a Q20 production floor, and 89.6% at L = 100 beyond the array's sensing horizon, where the faradaic axis resolves the purine pair the other two leave ambiguous. Because the readout is electrical and continuous, the throughput bottleneck moves off the fluidics and onto compute and array area, giving an areal base rate of ~2.1×10^6 bases/s per mm^2, under which a single-run genome becomes a parallelism condition (~2,200 mm^2 at 0.5 µm pitch) gated on Q20 accuracy rather than a time-to-result one. No chip exists yet, and every number here comes from a simulator. If the signal survives fabrication, what follows is a

cheap, reusable, fluidics-free sensor whose capability is set by the decoder: sequencing first on targeted panels, then the surface-binding and continuous-monitoring assays the read-length limit never touches, and eventually the population-scale sequence a general model of genotype to phenotype would need.

***Keywords*.** electrochemical impedance spectroscopy; label-free biosensing; sequencing by synthesis; passive microelectrode array; learned basecalling

I. Introduction

Suppose an event at an electrode leaves an electrical trace. A base added to a strand of DNA, a target binding to a probe, a molecule giving up an electron: each one moves charge, and moving charge changes how the interface looks to a small AC signal. If those changes are distinct enough to tell apart, then an array of electrodes plus a model trained to read them is a measuring instrument, and an unusual one. The electrode interprets nothing. It reports the state of its own surface. Whatever meaning that report carries is supplied afterward by the decoder, so the same chip can be pointed at a different question by retraining rather than refabricating, and it improves as models improve rather than as fabrication improves. Pitch and active-site size are chosen to match whatever is being measured; they are properties of the assay, not of the platform. This paper is about that instrument, developed against the hardest task we could give it.

We chose DNA sequencing as that task because the signal is weak, the event is fast, and the answer has to be one of four. If the architecture holds there, easier measurements follow. The reason to care is not the price of a genome, which has already fallen from roughly three billion dollars to a few hundred [1,2,3]. It is that most of the medicine waiting on sequence is waiting for it somewhere inconvenient. A model that learns how genotype produces phenotype needs sequence from millions of people, not thousands [4,5]. A clinician with a critically ill newborn needs an answer within the shift, not the week [6]. A diagnostic that belongs in a doctor's office needs an instrument that is small, cheap, and free of fluidics to maintain. None of these asks for a whole genome first. Each asks for a short, targeted read, produced cheaply, near the person who needs it.

Today's high-throughput platforms read a base one of two ways. They image a fluorescent label attached during each cycle, or they detect a chemical byproduct released when the base goes in [7,8]. Both work well, and both pay for it in the same currency. An optical platform needs something to attach a label to, an exposure to capture it, and a wash to clear the previous cycle before the next can start. The wash and the exposure set a floor on cycle time, demand fluidics precise enough to dominate the cost of the machine [1], and consume a flow cell that is discarded at the end of every run. Bulk electronic platforms drop the optics but still reset between cycles, and still read a chemical proxy rather than the incorporation itself [8]. In both cases the instrument spends most of its time on housekeeping: moving liquid, resetting the reaction, and taking a picture between every base of a process that is underneath continuous.

ECSEQ-0 does not label the base. It reads the electrical disturbance the incorporation itself makes. While a polymerase extends a surface-bound strand, the impedance of the electrode-electrolyte interface shifts, and the shift depends on which base went in. Measuring that shift by EIS at each pixel is the whole measurement. Nothing is labeled, nothing is imaged, and no reagent is exchanged between incorporations, because all four nucleotides are present throughout. What was a fluidics problem becomes a readout problem: how many pixels can be sampled, how fast, and how much compute the basecalling costs. That is a more tractable problem, because it is the kind that ordinary circuit engineering and falling compute prices keep solving.

Reading the event electrically also changes what else the chip can do. The array is passive and the discrimination lives in the model, so one array and one reader cover many assays with a retrained decoder. Sequencing is the first and hardest of them, not the whole of them, and it is the one that pays a specific price. The incorporation signal is visible only while the growing strand remains within a few nanometres of the electrode, which bounds reads to tens of bases. Sequencing therefore buys parallelism with read length, which is a fair trade, since time to a finished genome depends on read length only through how many runs it takes. Measurements of things that stay at the surface (a hybridized target, an antibody-antigen pair, a redox-active metabolite) never reach that limit at all.

Three things had to be true for this to be worth attempting now, and none of them held for earlier electrochemical sequencing work. Thin-film electrode arrays can be patterned densely enough to give each clonal cluster its own sensor at micron pitch. Polymerase kinetics

are understood quantitatively enough to connect an incorporation event to an expected electrical signature [9,10]. And models are now good enough to recover base identity from a noisy multi-dimensional spectrum by using the sequence around it. Each is unremarkable alone. Together they make per-event impedance sequencing worth testing.

Reading incorporation this way raises three questions, and they set the order of the paper. Is the signal actually base-specific and repeatable? We argue that it separates into three frequency bands with different physical origins: a kinetic band from how long the polymerase dwells, a molecular band from the nucleotide's dielectric coupling to the gold, and a faradaic band from base-dependent oxidation, which is what separates A from G. Together they carry enough information for all four bases. Is a single event strong enough to call? On its own, no, so each landed template is grown into a clonal cluster and read as an ensemble average, and the decoder reads a whole read in context rather than one event in isolation. Can the readout scale without an amplifier under every pixel, which would give back the cost advantage of a passive array? We treat that as a data-acquisition problem and work it through in the appendix.

The contributions are: (i) a passive electrochemical sensor architecture in which the same metal film anchors the DNA and transduces the incorporation signal; (ii) a physical account of the base-specific impedance signature and the three frequency-separated axes it decomposes into; (iii) a two-stage basecaller, a per-event spectral encoder followed by a temporal transformer that reads the cluster in context; (iv) an evaluation on physically simulated reads, with a throughput and read-length analysis that states the conditions under which the architecture reaches whole-genome scale; and (v) a front-end data-rate and basecalling-compute budget, together with an end-to-end simulation of the readout signal chain that locates its binding constraint at the analog front end. We keep read length and throughput apart throughout, because they are separate constraints: read length is bounded by interfacial screening geometry at the sensor, throughput by parallelism in the readout. Neither limits the other, and conflating them makes one read-length figure carry two unrelated claims. One caveat governs everything that follows. No chip exists yet. Every number reported here comes from a physics simulator, and fabricating the first pixel is our immediate concern after filing this whitepaper and the associated patents.

II. Related Work

Label-free and electronic sequencing. The closest commercial precedent is semiconductor sequencing, in which an array of ion-sensitive field-effect transistors detects the proton released upon nucleotide incorporation, achieving non-optical, natural-base sequencing at scale [8]. It reads a bulk pH change integrated over each well, however, and so does not resolve the incorporation event itself. Closer in architecture is molecular-electronics sensing, in which a passive semiconductor chip records single-molecule enzyme activity electrically in real time, including polymerase-class binding kinetics [11]. This is the nearest reported analogue to our present work, a passive electrode array reading an enzymatic event as it occurs, differing in that it resolves binding kinetics at individual molecules rather than an ensemble incorporation signature decoded into base identity. At the single-molecule limit, recognition tunneling through a functionalized nanogap established per-base electronic discrimination [12,13], and graphene field-effect transistors have since discriminated among individual nucleobases electronically [14]. Both operate on static molecules rather than during polymerase-mediated incorporation. ECSEQ-0 differs from all three in reading a per-event, per-pixel impedance perturbation generated during active synthesis.

Polymerase-kinetics readout. Single-molecule real-time sequencing demonstrated that polymerase dynamics are directly readable during synthesis, calling bases from the timing of fluorescent pulses as the enzyme incorporates them [9], and later that base modifications can be inferred from incorporation kinetics alone [15]. This grounds the premise underlying the kinetic axis that base-dependent polymerase dwell carries callable information. ECSEQ-0 reads the same kinetics through their effect on interfacial impedance rather than through an optical pulse.

Electrochemical impedance sensing of DNA. Electrochemical impedance sensing is an established, analyte-general biosensor modality on functionalized microelectrode arrays, applied label-free to proteins and pathogens as well as nucleic acids [16]. The platform's application-generality inherits from this biosensing lineage rather than from a sequencing-specific one, with sequencing its most demanding member. Within it, impedance-based DNA sensing is well characterized for hybridization,

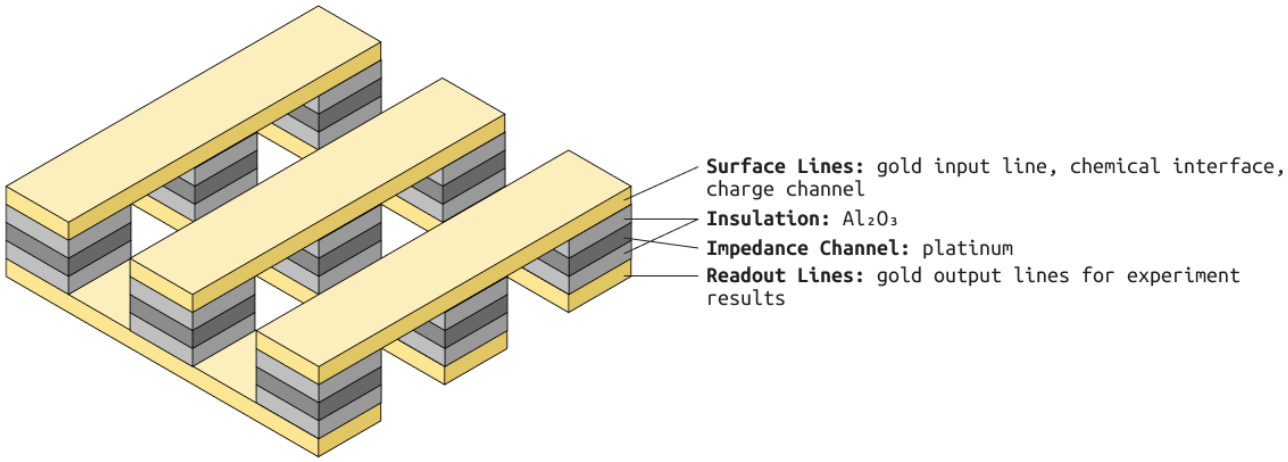

Fig. 2: Crossbar architecture. Gold surface lines carry the chemical interface and the charge channel; an Al_2O_3 layer isolates them from the Pt impedance channel; gold readout lines run beneath. One pixel is a single crossing, at 2 µm pitch with a ~1 µm active site.

where EIS detects the formation or mismatch of a preformed duplex on a functionalized electrode [17,18]. That work grounds the interfacial chemistry and measurement modality ECSEQ-0 builds on, but interrogates static duplexes rather than active synthesis. To our knowledge, no prior work demonstrates real-time base discrimination during synthesis by impedance, which is the capability we are targeting here.

III. Overview

ECSEQ-0 centers on a high-density array of passive electrodes, with each measuring the electrical impedance of the liquid immediately above it. A downstream model with per-task heads reads those measurements and infers what is happening there. Nothing in that chain is specific to DNA. What makes it a sequencer rather than a pathogen detector or a metabolite monitor is which decoder is loaded on top of it. This paper follows the chain end to end for sequencing, the hardest case we could find.

First a clonal cluster of identical template copies is grown over each pixel, lifting the per-event signal from a single molecule to a measurable ensemble average over the pixel's area. During synthesis, all four natural nucleotides are present, without wells, and a polymerase proceeds to extend each surface-bound primer. As each base is incorporated, an EIS sweep at that pixel captures the base-dependent perturbation of the interfacial impedance. This per-event spectrum is passed to a downstream basecaller. A per-cycle encoder compresses each spectrum into an embedding, and a temporal transformer reads the whole cluster's embedding sequence to call each base in sequential context (Fig. 1).

What bounds the method is the reach of the measurement itself. An electrode's field does not extend far into an electrolyte, because the dissolved ions rearrange to cancel it within a few nanometres. Each base incorporated carries the polymerase active site roughly a third of a nanometre further from the surface, so after some nine to twenty-six bases, depending on ionic strength, the event falls outside the region the electrode can sense. ECSEQ-0 is therefore a short-read instrument, and this follows from interfacial electrostatics rather than

from any limitation of the model. Resequencing against a known reference, targeted panels of defined genes, and most diagnostic questions are all answerable at this read length, and established short-read platforms sequence human genomes precisely this way, by covering each position many times over.

The sections that follow trace this chain in order. We first describe the sensor stack, the surface chemistry that makes an event measurable, the physical origin of the base-specific impedance signature, and the three axes along which it separates the bases, and then the model that decodes this signal. Because the platform is pre-hardware, the model is currently trained on reads from a physics-based simulator. We describe both the training procedure and the planned inversion to real data once a chip is available. We then report basecalling accuracy on simulated reads, and close with limitations and the path to hardware.

IV. Methods: Sensor Architecture and Signal

Sections A, B, and D describe the instrument itself: the electrode stack, its surface chemistry, and the interfacial-impedance signal it transduces along three discrimination axes. These are application-general. The same array and the same three-axis physics read any surface electrochemistry a decoder is trained to interpret, with pitch and active-site size chosen to match the analyte. Sections C and E are the sequencing-specific layer. Section F consolidates every assumption the whole architecture rests on into one auditable table.

A. Sensor Stack

Every pixel is a vertical stack of a top gold (Au) electrode and an underlying platinum (Pt) electrode, separated by anodized Al_2O_3 dielectrics (Fig. 2). The two metals serve different functions: the Au layer is the chemical interface that binds the DNA template, and the Pt layer is the impedance transducer. Using two metals rather than one is what makes frequency decomposition available. The Au and Pt interfaces relax on different timescales, so their contributions to the total impedance separate across frequency instead of summing into a single indistinguishable response.

The array is addressed passively over row and column interconnects, with a polymer fill isolating adjacent pixels to suppress crosstalk (Fig. 3). Localized per-pixel EIS is the measurement primitive throughout [19], and

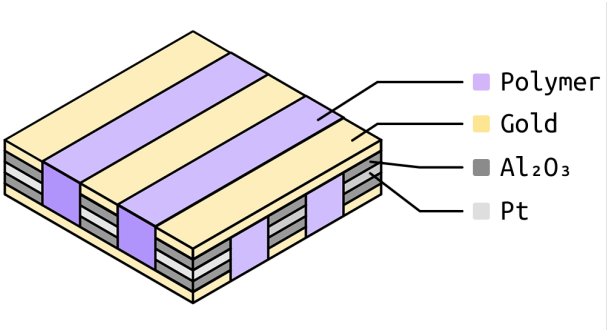

Fig. 3: 2 µm pixel pitch array architecture.

the Pt interface provides a stable, low-noise double-layer response across the full frequency sweep [20]. During operation, an EIS sweep at each pixel is synchronized to polymerase-mediated incorporation, and the resulting perturbation in interfacial impedance is the raw per-event signal passed to the basecaller. The off-chip readout that services the passive array, array substrate, excitation, analog front-end, code-division multiplex, and digital demux, is analyzed end to end as a five-stage chain in the appendix.

Device scales. The architecture is pitch-invariant with pixel spacing, so one design description covers devices of very different sizes. Three scales are referred to throughout. The prototype is a 4-pixel, direct-wired die which allows us to test the discriminating impedance signature. The full-spec die is a 960 × 960 array (921,600 pixels, 3.69 mm^2 active area), a discrimination demonstrator and targeted-panel device. Panel scale refers to a 2.25×10^8-pixel array (~900 mm^2), the regime in which whole-genome throughput comes within reach; closing a genome in a single run needs a larger array still. Sections 4 through 6 concern the architecture and are independent of scale. The throughput analysis in the appendix is a panel-scale projection and is marked as such. The three scales are separated by fabrication capabilities as well as by area. The prototype is patterned by single-field maskless UV exposure, within reach of a self-built exposure tool, whereas the full-spec die and panel scale require a foundry.

Operating point. These three scales are one axis of a larger operating point, scale (pixel count) × read length × application, on a single sensor architecture, not a family of platforms. The architecture is invariant; what moves along the continuum is where it is operated. At small array, short read, and known targets the same chip is a clas-

sification or targeted-panel device, and at panel scale with mappable reads and coverage it is a genome sequencer. Read length is not a free parameter: it is gated by the interfacial sensing envelope, short at the nominal operating point (~9 bases) and bounded above by convergent ceilings, so this remains a short-read architecture. Buying a long read back would mean reintroducing microfluidic control and a flow cell, which is precisely what the design exists to remove. Crossing from the near-term panel point to the genome point requires reads that clear the mapping floor, which is a two-component gate: long enough for unique placement (~20–25 bases) and accurate enough that an error-free seed survives (a lower per-cycle timing CV).

B. Surface Functionalization

The surface is functionalized with a self-assembled monolayer (SAM) of thiol-anchored primers, immobilized through Au-thiolate bonds [21]. The SAM does two jobs at once: it attaches the primers, and it is the dielectric layer whose perturbation the Au channel reads. Functionalization is therefore part of the transduction stack, not a substrate-preparation step that precedes it. Optical platforms do not face this constraint, because what the DNA binds to and what generates the signal are different objects: a substrate and a camera. Here they are the same layer, so every functionalization choice must satisfy two requirements simultaneously. The surface must present sufficient primer density to support bridge amplification, and it must not degrade the Au interface's own double-layer response, which is the baseline against which the molecular axis is measured.

Primer density and orientation. Three consequences follow from those requirements. First, orientation control is part of what makes the ~1 μm cluster geometrically reproducible from pixel to pixel. Second, density variance contributes to the ~25% pixel-to-pixel spread in charge-transfer resistance, which the model treats as a nuisance parameter. Third, the backfill layer belongs to the measurement rather than to sample preparation. Backfill is a short thiol-alcohol treatment that displaces nonspecifically adsorbed probe and sets hybridization-competent probe density [22], and because it also forms part of the baseline dielectric environment the Au interface presents, it is measured and held fixed by the blank-chip and post-functionalization EIS calibration steps. A change in backfill density is electrochemically indistinguishable from a change at the functionalization step, and the same calibration captures both.

Anchor chemistry and reuse. The 50 to 100-run reuse estimate is bounded by how many strip-and-reload cycles the thiol-Au bond survives before thiolate desorption and oxidation degrade it [21], not by anything about the DNA. Multidentate anchors, primers terminated in two or three thiol groups, chelate the gold at several points, so no single bond failure releases the strand, and trithiol-capped oligonucleotides survive thiol-competing conditions that strip monothiol-anchored strands from the same surface [23]. Whether this extends reuse at ECSEQ-0's operating conditions is untested, and the substitution remains a low-risk proposal pending hardware validation. Reuse is a cost lever essentially unique to ECSEQ-0. Competing platforms sequence on a single-use flow cell, chip, or wafer, so surface preparation is a per-run cost for them and a per-reuse-cycle cost here, and the reuse count multiplies directly into that advantage.

Between-run regeneration. Reuse requires a between-run cycle that removes the amplified DNA while leaving the thiol-Au anchor and backfill intact. A denaturing strip, such as a dilute NaOH or formamide wash, that clears duplex DNA without approaching the potentials or temperatures that desorb thiolate. Regeneration of this kind is standard across biosensor platforms [24]. The strip resets the surface to its pre-loading state, re-verified by the same blank-chip and post-functionalization EIS pair used at first build. Protocol details, reagent, exposure time, and the verification threshold for reload, are engineering parameters to be established on hardware.

Physical confinement. Lateral isolation between pixels, the polymer insulating fill, plays the role a patterned nanowell or charged capture pad plays on competing platforms. It converts amplification from a random-deposition process into one with an addressable site per cluster. High-density clonal arrays increasingly favor geometrically predefined single-cluster sites over random deposition, because random deposition cannot suppress the co-localization failure mode (two templates merging into one unresolvable signal) at high density [1,25]. ECSEQ-0's version of this constraint is defined by the SAM/backfill boundary at the pixel edge rather than a physical well wall. A second and distinct failure mode follows from that choice. Co-localization concerns two independently landed templates merging; overgrowth con-

cerns a single, correctly seeded cluster whose own bridge amplification carries it past its allotted area. At a 2 µm pitch with a ~1 µm active site, a well-formed ~1 µm cluster reaches 0.5 µm from its pixel centre and the neighbouring sensing zone begins at 1.5 µm, so the cluster has to roughly triple its radius before it starts reporting into its neighbour. That margin is set by amplification cycle count and kinetics rather than by landing statistics, and without a physical wall to arrest growth there is nothing but reaction time bounding it. No bound on cluster radius against cycle count is established here; it is a single-pixel or bulk-surface imaging measurement, independent of electrochemical readout, and it is a validation target.

C. Preparation and Loading

Sample preparation uses conventional techniques of short-read platforms. Genomic DNA is fragmented to the target read length, and adapters complementary to the two surface primers are ligated to the fragment ends so each strand can hybridize to the primer lawn. This is a borrowed, well-characterized primitive rather than a novel step. As the throughput analysis in the appendix shows, however, its wall-clock cost dominates the per-run time budget. Whenever a configuration closes a genome in one run ($\eta PL \geq B$), sequencing occupies well under 1% of wall-clock time. At $L = 30$ roughly 3.5 seconds against a ~35-minute prep, so $T_{\text{genome}} \approx T_{\text{prep}}$ and prep is not merely the single largest term in time-to-genome but effectively the only one, even though it is not where the platform's technical differentiation lies. Prep chemistry is not a research question for ECSEQ-0, but prep time is therefore the single lever on time-to-genome, and the subsections below treat fragmentation, loading, amplification, and quality control as separable levers.

Fragmentation and Adapter Attachment. Fragmentation and adapter attachment may proceed by either of two routes. The conventional route employs mechanical or enzymatic fragmentation followed by separate end-repair, A-tailing, and ligation. It is the most widely supported and the slowest. Tagmentation collapses both operations into a single step using a hyperactive Tn5 transposase [26], and rapid transposase kits report hands-on times of approximately ten minutes rather than hours [27]. Tagmentation would therefore reduce prep time substantially while leaving the array surface unaffected, as both routes yield a standard adapter-bearing fragment. The selection is accordingly a bench decision independent of the sensor, and constitutes the most concrete and lowest-risk lever on the prep-time budget, pending a direct comparison of the two routes against ECSEQ-0's design.

Loading. The prepared library is flowed over the functionalized array under conditions favouring sparse, single-fragment landing at each pixel. With Poisson loading the library is diluted so that most sites capture zero or one molecule, with empty and multiply-occupied sites discarded downstream. Poisson statistics cap single-occupancy near 37% whenever discrete sites are seeded at random [25,28], and competing platforms either accept that ceiling or engineer past it. Some patterned flow cells race amplification kinetics against molecule arrival to suppress multi-templating at a fixed site. Nanoball platforms capture pre-formed amplicons onto lithographically patterned, charged binding pads sized for a single nanoball, and several bead platforms avoid on-array loading statistics altogether, amplifying clonally off-array by emulsion PCR onto a bead and loading only well-amplified particles [25,29,30]. Off-array amplification is structurally incompatible with the sensing mechanism, which requires the polymerase and its growing strand to sit directly at the electrode stack, a pre-amplified bead would still need that contact, adding a capture problem rather than removing a step. Kinetic-exclusion loading is not obviously incompatible with ECSEQ-0's chemistry, but it is actively defended elsewhere in the field.

ECSEQ-0 has a loading-control option unavailable to the surveyed platforms, because its array is already a grid of individually addressable electrodes before sequencing begins, and the row and column addressing built for EIS readout can in principle be repurposed during loading. DNA is strongly polyanionic, carrying roughly one negative charge per base along the phosphate backbone. A modest, transient positive bias at a target pixel should therefore raise the local electrophoretic or dielectrophoretic capture rate of nearby template strands relative to unbiased neighbours, biasing single-molecule seeding toward the addressed site without racing amplification kinetics or pre-sorting particles off-array. The hardware required is already present for readout, so this would be close to a zero-incremental-component loading control, unlike patterned charge deposition (a separate fabrication step) or bead sorting (a separate fluidic subsystem). Active-bias loading control is a distinct mechanism from both kinetic-exclusion loading and patterned-charge-pad capture, and is gated on single-pixel hardware validation.

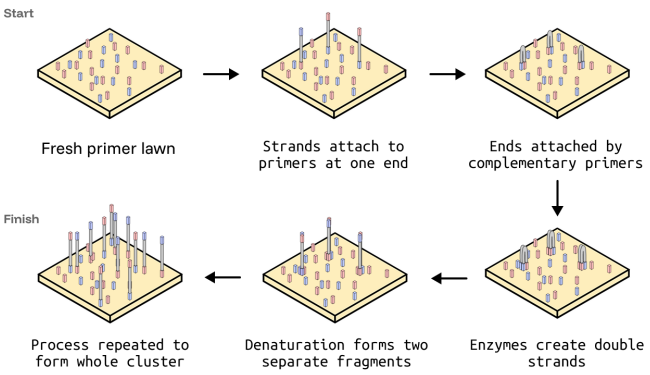

Fig. 4: Isothermal bridge amplification grows a clonal cluster in place from the primer lawn. A landed template attaches to a surface primer, its free end bridges to an adjacent complementary primer, the bridge is extended and denatured into two surface-bound copies, and the cycle repeats to build the full ~1,000-copy cluster.

Clonal Amplification. A single template does not produce enough signal to measure, so each landed fragment is copied in place on the primer lawn by bridge amplification [31]. The free end of a surface-bound strand hybridizes to an adjacent complementary primer, a polymerase extends the bridge, and denaturation releases two surface-bound copies. Iterating this cycle grows a dense clonal cluster of roughly a thousand identical copies over a single pixel (Fig. 4), lifting the per-event signal from a single molecule to a measurable ensemble average. Amplification runs isothermally (Bst 2.0 WarmStart, ~65 °C), at the same operating temperature as sequencing, rather than the thermal-cycled PCR that emulsion-based platforms require off-array. Bridge amplification was chosen over rolling-circle amplification because this requires the template to be circularized first and produces a concatemer that must then be captured onto a patterned site as a separate step. Bridge amplification instead grows the cluster outward from the same primer lawn already required at the electrode for DNA anchoring and molecular-axis sensing, so the cluster is never detached from the electrode.

Occupancy and Quality Control. Polyclonal pixels, carrying more than one template, and under-loaded pixels, empty or with failed amplification, must both be identified and excluded before or during sequencing. Optical platforms detect polyclonal sites by imaging, since an overlapping or double-intensity cluster is visually distinguishable from a clean single cluster, but this requires the image in the first place. ECSEQ-0 obtains an equivalent diagnostic from a step the calibration chain already performs. The blank-chip and post-functionalization EIS baseline established for every pixel before sequencing sets an expected impedance signature for a well-formed ~1,000-copy cluster. A pixel falling well outside that range is flagged without a separate imaging subsystem.

D. Electrochemical Signature

Nucleotide incorporation produces localized, base-dependent perturbations in the interfacial impedance. Modelling the electrode-electrolyte interface as a series of constant-phase-element double layers coupled to a charge-transfer pathway, we write the total impedance as the sum of the Pt and Au contributions, $Z(\omega) = Z_{\mathrm{Pt}}(\omega) + Z_{\mathrm{Au}}(\omega)$. Because the two interfaces have distinct electrochemical relaxation time constants, their contributions decouple across frequency bands, encoding base identity along three physically distinct axes.

Kinetic axis. The kinetic axis dominates the lower-frequency regime, 100 Hz to 10 kHz, where the Pt double-layer response is strongest [20]. During incorporation the polymerase passes through an open nucleotide-recognition state and a closed ternary conformation, the latter established as the induced-fit step that discriminates correct from incorrect nucleotides [10]. Single-molecule electrical recording of a polymerase resolves the two states separately and places the base-dependence in the open one: the closed complex lasts 0.3–0.4 ms and is largely independent of which nucleotide is incorporated, while the open state runs 38–71 ms and differs about two-fold between A·T/T·A and G·C/C·G pairs [60]. The base-dependent quantity is therefore the cycle duration rather than the closed-fingers residence time, and because the perturbation persists for that duration it modulates the response in this band. Individual dwell times vary stochastically [9,60], but aggregating over a cluster recovers sequence-dependent temporal patterns usable as classification features.

Molecular axis. The molecular axis appears in the high-frequency tail of the sweep, 10 kHz to 100 kHz, where the Au interface response is most pronounced. It reflects the dielectric and dipole coupling of the incorporated nucleotide to the thiol-SAM interface, a base-dependent perturbation of order 15% in interfacial impedance magnitude. Unlike the kinetic axis it is dwell-independent, re-

lying on base-intrinsic amplitude, and so supplies discrimination orthogonal to dwell. The mechanism is well supported at the molecular end and only partly at the device end: the four bases differ measurably in dipole moment and polarizability, with computed polarizabilities ordering C < T < A < G [32], DNA bound at a gold electrode measurably alters that electrode's interfacial impedance [33], and per-pixel interfacial-impedance readout has been demonstrated on a CMOS nanocapacitor array architecturally close to ECSEQ-0's own, though on microparticles and living cells rather than on DNA [34], with high-frequency operation reported to sense past the screening length under physiological salt on a different transducer [35]. The array and the readout modality therefore have precedent; per-base dielectric discrimination on them does not. That array work operates up to roughly 50 MHz [34], so what remains to be established on hardware is both the coupling itself and the band in which it is best resolved. The assignment to 10-100 kHz is the component of the specification most likely to move, and the encoder's molecular branch is given the full sweep rather than the high tail alone so that it can locate the perturbation rather than assume it.

Faradaic axis. The faradaic axis resolves residual ambiguity within the purine pair, exploiting the distinct intrinsic oxidation potentials of the natural nucleotides. Guanine has the lowest oxidation potential, followed by adenine, with the pyrimidines requiring higher and more sluggish potentials [36]. The same ordering holds for direct oxidation at an electrode, where guanine and adenine oxidize by mechanistically distinct, well-separated voltammetric steps [37]. Biasing the Au electrode toward the guanine oxidation onset therefore renders charge-transfer resistance base-dependent, and this signal is assigned to the low-frequency shoulder of the spectrum, 20 to 90 Hz.

The band assignment is a modelling assumption rather than a measured result. The charge-transfer time constant is system-specific, so no universal band exists to assign it to, and published DNA-on-gold impedance work sweeps broadly without delineating one [38]. That work also relies on a solution redox mediator, whereas the faradaic axis proposes direct oxidation of the bases free of mediators, conventionally characterized by DC voltammetry rather than AC impedance [37]. The simulator's own diagnostic sharpens the caveat rather than settling it. Isolating the faradaic R_{ct} contribution from the dielectric response already present on the Au channel gives a per-event discriminability of $d' \approx 1.5$ against the dielectric effect's ≈ 2.1, and the isolated faradaic signal is about as visible in the main 100 Hz–100 kHz sweep as in the 20–90 Hz sub-band. Two things follow. The sub-band is where the arc is expected to sit at the modelled parameters, not a window the signal is confined to, and the axis is the weaker of the two Au-channel effects on a single event. Its value is that it separates the purines, which the stronger dielectric effect does not, rather than that it is the larger signal. A charge-transfer resistance closer to the SAM-blocked baseline would push the arc down into a band the main sweep cannot reach and make the sub-band exclusive, which is a design parameter rather than a fixed property.

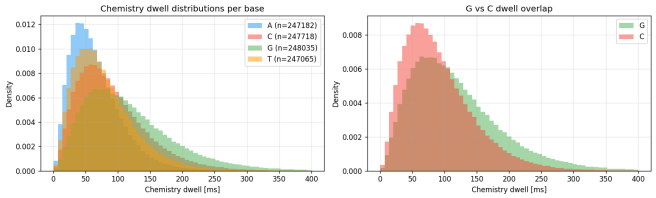

Fig. 5: Simulated per-base dwell distributions overlap heavily; dwell alone cannot separate the bases.

Single-pixel validation therefore targets two quantities: the band in which the arc is accessible at ECSEQ-0's geometry, and the bias-control tolerance compatible with polymerase fidelity and template integrity. Polymerase kinetics set the frequency floor so the measurement reads the high-frequency shoulder of the charge-transfer arc.

E. Sequencing Loop

Sequencing proceeds with all four natural nucleotides co-present in a single pool over the surface-bound clusters, generated beforehand on the primer lawn by bridge amplification [31], run isothermally here with Bst 2.0 WarmStart at ~65 °C rather than under the thermal cycling of the original protocol. A polymerase extends each surface-bound primer processively, incorporating the complementary base at each position, and that incorporation is read in real time from the impedance signature of the event. In conventional cyclic sequencing the wash clears unincorporated labeled nucleotides before readout and resets the reaction between cycles [7]; reading the intrinsic electrochemical signal removes the need for it, since there is no label to image, nothing to cleave, and no reagent to exchange between incorporations. Wash-free operation follows from label-free real-time transduction rather than being engineered separately.

This places ECSEQ-0 among a small group. Most es-

tablished platforms are structurally cyclic, repeating add-reagent, detect, clear-and-reset once per base or per flow even where an upstream step is itself isothermal [8,29,39]. Only two surveyed architectures are continuous: real-time single-molecule optical sequencing, which reads fluorescence optically from one confined polymerase with no wash at any point in a run [9], and native nanopore sequencing, which reads ionic current as a strand threads a pore but is translocation rather than synthesis and so has no incorporation event to read [40,41]. ECSEQ-0 is closest to the former, both continuous, real-time, synthesis-driven, and wash-free, differing in that it reads an ensemble of roughly 1,000 co-synthesizing copies at one electrode rather than a single polymerase inside a nanophotonic structure. No platform surveyed combines continuous synthesis with a fully electrical, label-free readout and clonal-ensemble amplification. Scaling by pixel count rather than optical field of view, it offers potentially higher throughput than either.

Homopolymer runs. A repeated base is where the competing electrical and flow-based chemistries struggle. Flow-based semiconductor sequencing dispenses one nucleotide species at a time and infers run length from the magnitude of a single summed signal, so distinguishing five identical bases from six means resolving a noisy continuous magnitude into an integer. Independent benchmarking identifies this as that platform's characteristic error mode, with miscalls rising sharply as runs lengthen [42,43]. Nanopore current is ambiguous for a different reason. Several bases occupy the constriction at once, so the current reports a k-mer rather than a single base [44]; within a homopolymer successive k-mers are identical, so the signal barely changes [45]. At least one nanopore chemistry exists specifically to convert each base into its own coded reporter, so a homopolymer reads as separable events rather than one smeared excursion [46].

ECSEQ-0 was never flow-limited. All four nucleotides are continuously present, so each incorporation is a separate kinetic event, spaced from its neighbours by the polymerase's own kinetics rather than by an imposed flow boundary. If successive events stay resolvable at the sensor's acquisition rate, a homopolymer should read as a train of dwell-resolved impedance events rather than one summed magnitude. Laminar shear across the array, holding the duplex near-parallel to the surface, is one long-term route to separating those events spatially as well as in time. It stands as a testable prediction until dwell-resolved data over a repeated-base template is in hand.

Desynchronization without a cycle boundary. Cyclic optical platforms lose per-base quality over a long read, with phasing and prephasing well-characterized contributors alongside signal decay and optical cross-talk [47,48,49,50]. At each cycle boundary some strands in a cluster fall behind through incomplete extension, and others run ahead through incomplete terminator or dye removal. Because every cluster is locked to the same clock, the error compounds cycle over cycle. Measured rates are 0.5 to 0.8% phasing and 0.4 to 0.6% prephasing per cycle [48].

ECSEQ-0 has no such clock. Each of the ~1,000 copies in a cluster synthesizes continuously and independently, with no wash or cleave step all copies must clear together. Stochastic per-base kinetics still spread incorporation events apart over a read, which the simulator models as an independent noise source. The consequence is the same: read length is bounded by how far the copies drift apart. Losing the clock is not a win. It trades a compounding error for an unbounded random walk, and the walk is worse: spread grows as $0.485\sqrt{n}$ bases and passes one full base within about four cycles, roughly two orders of magnitude beyond what a cyclic platform contends with. Wash-free operation buys a simple instrument and pays for it with a short read.

Reagent Supply Over the Run. With no per-cycle wash, nucleotides are consumed and byproducts, pyrophosphate and protons, accumulate continuously rather than being cleared each base. Continuous real-time optical sequencing runs a single loaded chamber to completion without reagent exchange, an option available to it because each confinement site observes a tiny volume and synthesizes correspondingly little DNA [9]. ECSEQ-0's ensemble clusters consume far more nucleotides per pixel, across ~1,000 copies per cluster and up to the full array. Whether the bulk pool needs periodic replenishment or can run from a single loaded volume is an open engineering parameter. Either way it does not reintroduce a fluid exchange step. A bulk exchange on the timescale of a whole run is a different operation, with different fluidic requirements, from the per-base exchange a cyclic platform performs hundreds of times.

Faradaic Interrogation. The faradaic axis runs by default, which raises one scheduling choice. The elevated-potential guanine interrogation can run continuously

alongside the low-potential kinetic and molecular sweep, or as short pulses interleaved between longer low-potential periods. Pulsing gives up some faradaic duty cycle, and so some fraction of events that benefit from the third axis, in exchange for a bounded exposure budget per run rather than an unbounded one. Which schedule is required, and what duty cycle is safe, is among the questions single-pixel validation is designed to answer.

F. Assumption Register

The architecture rests on a set of modelling assumptions and unvalidated mechanisms. Each is flagged where it arises, here or in the appendix; Table 1 consolidates them so they can be audited and, more to the point, counted. Collecting them hides nothing. It does the opposite: it makes the open surface of the design finite and visible on one page, with each entry's status and the measurement that would settle it.

Two things are worth reading off the table. The entries are concentrated in the surface chemistry and the readout back end rather than in the sensing thesis itself, and most of them are settled by the first two pieces of hardware, a single pixel and then a small array. Every result in this paper is conditional on this register, and its most load-bearing entry is the first line of the next experiment: that a real incorporation produces a base-discriminating impedance signal at all.

V. Model

The basecaller has two stages: a per-cycle encoder that compresses each spectrum into an embedding vector, and a temporal transformer that reads the cluster's full embedding sequence and predicts each base in context (Fig. 6). The split follows the structure of the problem. A single incorporation event is ambiguous, so the encoder does not call the base but organizes a separable representation of it, and the temporal stage resolves the remaining ambiguity by reading each event alongside its neighbours. The array itself is not specific to sequencing: it reports a per-pixel impedance spectrum and the interpretation lives entirely in the model, so the same hardware with a retrained decoder could in principle read other surface processes.

A. Input Representation

Each pixel reports one complex impedance spectrum per incorporation event: a 62-point frequency sweep, comprising a 50-point main band from 100 Hz to 100 kHz and a 12-point faradaic sub-band from ~20 to 90 Hz.

Perturbation representation. The model consumes not the raw spectrum but the change between two states on the same pixel, the impedance during an incorporation against that pixel's quiescent baseline, in polar form, as a log-magnitude change and a phase change at each frequency, concatenated into a 124-dimensional vector (Fig. 7, Fig. 8).

$$\Delta \log |Z| = \log\left(\frac{|Z|_{\text{event}}}{|Z|_{\text{baseline}}}\right) \tag{1a}$$

$$\Delta\varphi = \varphi_{\text{event}} - \varphi_{\text{baseline}}. \tag{1b}$$

Two physical reasons motivate this. The log-magnitude difference makes the input a ratio against the pixel's own baseline, which cancels per-pixel offset. Absolute impedance varies substantially across the array, with charge-transfer resistance spanning roughly 25% pixel to pixel from fabrication and functionalization tolerances, but the event-to-baseline ratio within a pixel is stable regardless, so the encoder never has to learn around a scale it cannot see at inference. Phase is orthogonal to magnitude, capturing the shift in reactive/resistive balance that magnitude alone misses, and together the two form the natural polar description of an EIS perturbation.

Dwell time. Dwell time is carried alongside the spectrum as a separate scalar rather than folded into it, because dwell is a distinct physical quantity, the base-dependent duration of the polymerase's catalytic cycle, dominated by its open-state recognition step [60], not a spectral feature. It is the physical carrier of the kinetic axis, and the quantity the encoder would otherwise have to recover indirectly from the low-frequency band. Dwell distributions are strongly right-skewed and overlap heavily between bases (Fig. 5), so the scalar is log-normalized against a 100 ms reference, which compresses the tail and lets the network see a signed deviation from typical dwell rather than an absolute count. It is supplied a second time to the temporal stage as a per-position bias, because the encoder's fused embedding discards most fine-grained dwell statistics.

B. Encoding Stage

Branches. The first stage is a tri-branch 1D convolutional network whose branches mirror the three discrim-

Table 1: Assumption register. Every modelling assumption, proposal, and unquantified mechanism the architecture depends on, with its status and the measurement that retires it. Status is one of *hypothesis*, the claim the platform rests on; *assumed*, a value, band, or protocol taken as given; *proposed*, a mechanism asserted but never tested; *unquantified*, known to matter with no bound established; *simulated*, supported by simulation alone; and *contested*, measured evidence disagrees with what is used here.

Assumption or mechanism	Status	What would settle it	Where
Incorporation produces a reproducible, base-discriminating impedance signature at a ~1 μm^2 pixel	hypothesis	single-pixel EIS during synthesis	§IV.D, §VII
Molecular axis: per-base dielectric coupling resolvable in the 10–100 kHz band at a gold electrode	assumed	single-pixel EIS on a functionalized Au electrode	§IV.D
Faradaic axis: signal in the 20–90 Hz band, and elevated Au bias compatible with polymerase fidelity	assumed	single-pixel faradaic interrogation	§IV.D, App. D
Faradaic exposure: irreversible guanine oxidation dilutes the newest event against prior bases still inside the envelope	simulated	duty-cycle comparison, continuous against pulsed bias	§IV.E, App. E.3
Faradaic axis's contribution to accuracy	unquantified	a matched faradaic-on/off ablation carried to completion at one read length	§VI.B
Surface reuse of 50–100 runs	assumed	strip-and-reload cycling on hardware	§IV.B
Multidentate anchor extends reuse past the monothiol baseline	proposed	qualification against the monothiol baseline	§IV.B
SAM-preserving denaturing regeneration strip	assumed	characterize the strip on hardware	§IV.B
Cluster does not overgrow its pixel within the amplification cycle count	unquantified	cluster radius against cycle count, by imaging	§IV.B
Active-bias loading control biases single-molecule seeding	proposed	capture rate at the pixel geometry	§IV.C
Occupancy QC against the pre-sequencing blank-chip and post-functionalization baseline	assumed	test against known multi-templated pixels	§IV.C
Homopolymer runs resolve as dwell-separated event trains	proposed	dwell-resolved data over a repeated-base template	§IV.E
Cluster stays in positional register long enough for adjacent incorporations to resolve	simulated	per-copy cycle-time distribution for Bst 2.0, specifically its CV	App. F.3
Passive-matrix line RC does not derate the 100 kHz band at array scale	unquantified	line-RC modelling, then measurement	App. E.4
Shared-converter readout preserves the basecall	simulated	front-end noise at the real integration time	App. E.7
Analog baseline nuller holds drift to $\leq$0.52% between calibration and event	simulated	nuller drift on hardware	App. E.7
Sensing horizon extendable from ~9 bases to the ~20–25 base mapping floor	unquantified	screening extension over a catalytically active polymerase	App. F.2
Simulator parameters (dephasing rate, band assignments, noise magnitudes) partly estimated	assumed	recalibration against a measured chip	§VI
Per-base dwell ordering (G slowest, A fastest)	contested	per-base cycle-time measurement for Bst 2.0 at 65 °C; the nearest single-molecule result groups by pair type instead	§IV.D, §VI

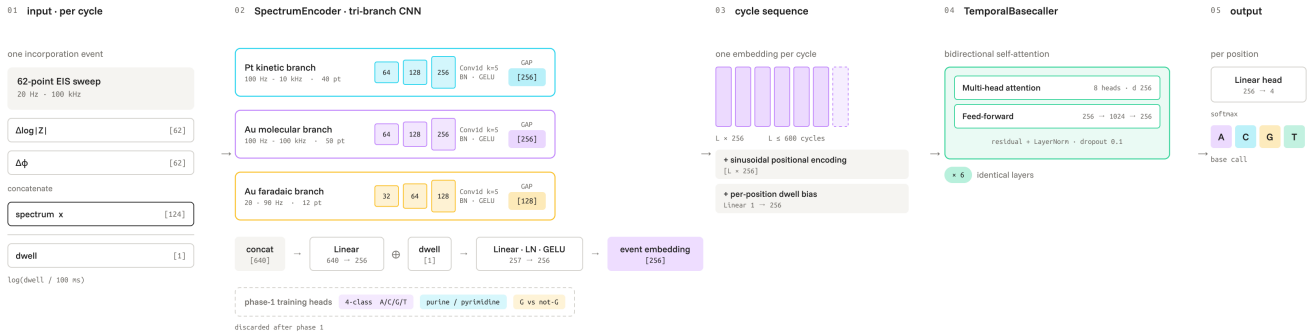

Fig. 6: The two-stage basecaller: a tri-branch per-event encoder feeds a temporal transformer that calls each base in sequence context.

ination axes. A kinetic branch convolves over the low-frequency band, roughly 100 Hz to 10 kHz, carrying the Pt dwell-correlated signal. A molecular branch spans the full 100 Hz to 100 kHz main sweep, carrying the Au dielectric and dipole signal. A faradaic branch reads the ~20 to 90 Hz sub-band, carrying the base-dependent charge-transfer signal that separates A from G. Making the decomposition explicit gives each branch a well-scoped target rather than forcing one network to disentangle the axes from the full spectrum. The faradaic branch's

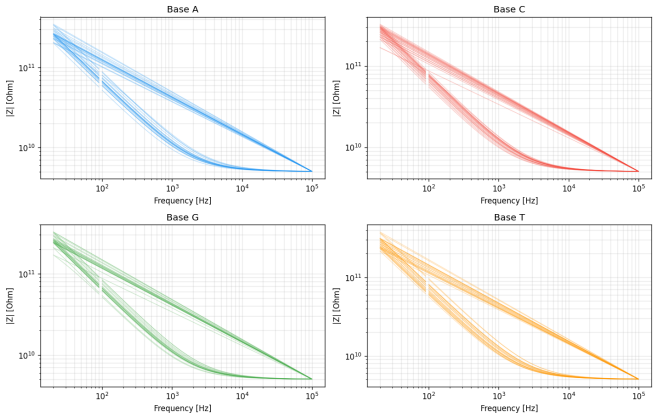

Fig. 7: Raw baseline and event spectra, one event per base. They are nearly identical: absolute |Z| carries scale, not base identity.

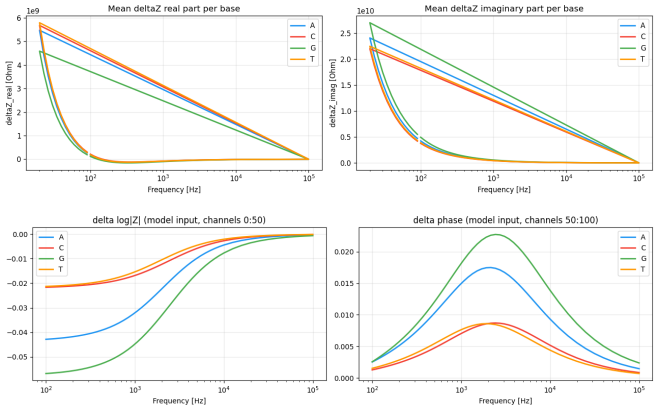

Fig. 8: The $\Delta\log|Z|$ and $\Delta\varphi$ perturbation actually fed to the encoder for the same events. This is the signal the model sees.

band is a scoping choice on the same footing: in simulation the same charge-transfer signal is about equally present in the main sweep (§IV.D), so the sub-band is where the branch is pointed rather than the only place the signal lives. The molecular branch is given the whole main sweep rather than only the 10 to 100 kHz tail where the Au response physically dominates, so that it can reference the dielectric perturbation against the lower-band baseline instead of assuming the tail's location in advance.

Layers. The kinetic and molecular branches are convolution stacks of channel width 64, 128, 256, with GELU activations, batch normalization, and light dropout. The faradaic branch is a lighter stack at 32, 64, 128. Each is followed by global average pooling. The three pooled outputs are concatenated to 640 dimensions and projected to a 256-dimensional embedding, onto which the log-normalized dwell scalar is concatenated and projected back to 256.

Training. The encoder is trained as a per-event classifier and reaches ~50% accuracy on the current L=30 reads. That is far short of production grade, because a single event remains ambiguous. Its purpose is to structure a separable embedding space rather than to call bases. Three heads run during pre-training. A primary 4-class head calls A/C/G/T. A purine-versus-pyrimidine head organizes the embedding along the axis the physics separates most cleanly, so the temporal stage inherits a representation in which the within-group pairs A/G and C/T are already partly separated. A G-versus-not-G head gives the faradaic branch a direct gradient toward the guanine-oxidation feature the other two are blind to. Label smoothing softens the two confusable pairs during training. The faradaic branch supplies the per-event separation of the A/G pair the other axes lack, and guanine in particular is resolved cleanly. A completed faradaic-on versus faradaic-off ablation is outstanding at both read lengths.

C. Temporal Stage

Architecture. The next stage is a transformer encoder over the sequence of per-cycle embeddings, on the order of 6 layers and 8 attention heads, with sinusoidal positional encoding and pre-layer-norm blocks. Because basecalling is offline, attention is bidirectional, so each position reads both upstream and downstream context. Dwell is injected again here as a learned per-position bias, because the encoder's fused embedding discards most fine-grained dwell statistics. The context window holds ~600 cycles. This is an architectural capacity, not a claim about read length, which the sensing geometry bounds to tens of bases. The window is sized well above that bound so the same model serves whatever length the sensing geometry permits.

Output. A linear head maps each position to four base logits, one set per pre-segmented event. The stage emits calibrated per-base posteriors rather than bare argmax calls. Every downstream variant caller consumes per-base quality and weights by it, so a calibrated softmax distribution over {A,C,G,T} is a required output of the architecture, not an optional addition.

Performance. This stage carries most of the remaining accuracy, resolving ambiguity single events cannot by

reading each base alongside its neighbours' kinetics and local sequence. It lifts the call from the encoder's ~50% per-event to 71.2% per-base, Phred Q5.4, on the current L=30 reads. That is a substantial gain but still short of the Q20 (99%) production floor. The residual and the path to closing it are set out below. On a production free-running stream this stage becomes a segmenting decoder over frames rather than a per-position classifier over events.

D. Training

Data. The model is currently trained on synthetic reads from a physics-based simulator, which produces ground-truth-labelled spectra by modelling the full chain from polymerase kinetics through Pt/Au transduction, cluster dephasing, and correlated noise.

Phases. Training proceeds in stages rather than end-to-end from scratch, which stabilizes the two differently shaped stages and keeps cost down. The encoder is pre-trained first on individual events under the multi-task objective above. Its weights are then frozen and its embeddings pre-computed once over the full dataset, so the temporal transformer trains against fixed embeddings without re-running the encoder at each step. A short lower-learning-rate phase fine-tunes the temporal stage. An optional final phase unfreezes the encoder for end-to-end fine-tuning at a much smaller encoder learning rate, protecting the pre-trained features while letting the two stages co-adapt. Training runs in bfloat16 under AdamW throughout.

Calibration. Because the model must emit usable quality scores, calibration is a training-time objective alongside accuracy, evaluated by expected calibration error (ECE) and a reliability diagram rather than by accuracy alone. On the current reads the raw softmax is already well calibrated at ECE 0.045, and temperature scaling does not improve it, so the raw posteriors are used directly as per-base quality scores.

Post-fabrication. The synthetic-first procedure then inverts. The simulator is recalibrated against the measured parameters of the real chip, and the model is fine-tuned from its existing checkpoint onto real data rather than re-trained from scratch. A learned frequency-band selector, a differentiable mask that collapses the 62-point sweep to only the frequencies carrying signal [51,52], would both speed hardware acquisition and serve as a separability diagnostic.

VI. Experiments

Every result below is measured on simulated data, not on hardware. The simulator generates ground-truth-labelled spectra by modelling the full chain from polymerase kinetics through Pt/Au transduction, cluster dephasing, and correlated noise, so what follows reports the decoder's ability to invert that simulated physics rather than a measurement of a device. These numbers indicate the signal carries the information the model needs to separate four bases, and that a decoder can recover it, which is the narrower question that has to be answered first: an instrument whose signal cannot be decoded in principle is not worth fabricating. What follows characterizes the accuracy the two-stage decoder reaches, and where its residual error falls against where the physics predicts it should.

Simulator provenance. Most of the simulator's parameters are literature-derived, including polymerase dwell distributions and base ordering, scaled from single-molecule real-time interpulse durations [9], base dielectric and oxidation properties [32,36,37], and the CPE-Randles interface model. The per-base ordering is the weakest of these. Its two-fold spread agrees with direct single-molecule measurement, but that measurement groups the bases by pair type, A·T/T·A slow against G·C/C·G fast, rather than reproducing the ordering used here, and one smFRET study reports no significant base dependence at all [60]. The cluster-dephasing rate, the band assignments for the molecular and faradaic axes, and the correlated-noise magnitudes are estimated. How many there are matters more than any single value, since together they set how far these results rest on the simulator rather than on measured physics. Each is retired at post-fabrication recalibration, when the simulator is fit to the measured chip. We report two training runs on the same architecture at two read lengths. The short-read run (L=30) sits at the top of the range the sensing geometry could permit, marginally past the 26-base ceiling the most dilute polymerase-compatible buffer allows, and is the operating point of targeted panels and resequencing. The long-read run (L=100) sits beyond that horizon, so it measures decoder capacity given a long context rather than a supported operating point. We report it because it bounds what the temporal stage can do given context the sensor cannot yet supply.

A. *Short-Read Run ($L = 30$)*

We evaluate the full pipeline on held-out simulated reads with the faradaic axis enabled. Accuracy is reported per-event, the encoder calling each incorporation in isolation, and per-base, the temporal stage calling each position in cluster context. Both are four-class problems at 25% chance. The per-event encoder reaches 50.1%, as expected for a single ambiguous spectrum read without temporal context. No faradaic-on/off ablation has been completed at either read length, but the per-class results below show the faradaic axis working, with the guanine and A/G channels essentially resolved. The temporal stage carries the call the rest of the way, lifting per-base accuracy to 71.2%, or Phred Q5.4 under $Q = -10\log_{10}(1-\text{acc})$. Unfreezing the encoder for a final phase changes this negligibly, from 71.0% to 71.2% (Fig. 9), indicating that the frozen-encoder representation is already near the joint optimum.

Q20 (99%) is the production floor, with Q30 (99.9%) the target for consensus over coverage, and the current figure sits below both. Two things indicate the gap is addressable. The residual is structured rather than random, concentrating on a single base pair rather than spreading across all four, which localizes it to a missing discrimination axis rather than to noise. There is also direct precedent on the decoder side, since successive basecaller architectures on nanopore raised both read and consensus accuracy substantially on an unchanged physical signal, with network capacity and better-matched training data each contributing measurable gains [53,54]. Together they place the shortfall in the current decoder and signal model rather than in the measurement itself.

The residual error is not distributed as the earlier synchronous-ensemble model suggested. The faradaic axis performs as intended: guanine is called at 98.8% with no dominant confusion, and the A↔G purine pair is essentially resolved, accounting for 1.3% of all errors and under 0.4% of all calls. The limiting channel is instead the pyrimidine pair, which the faradaic axis does not target by design. Adenine is called at 75.3%, losing predominantly to thymine, and cytosine at 65.2%, losing to thymine and adenine; thymine itself is called correctly in only 45.0% of cases, and C↔T constitutes the largest single error channel (Fig. 10). Errors are consequently divided approximately evenly between transitions (46.7%) and transversions (53.3%), so the residual is not transition-shaped. It is systematic, attributable to one base, one interface, and one absent discrimination axis, and therefore recurs across independent reads at the same reference position rather than averaging out with depth. The indicated mitigation is a fourth discrimination axis, or stronger dielectric contrast, for the C/T pair.

Table 2: Probability that a read is error-free, acc^L, at the current accuracy and at the Q20 and Q30 targets. The $L = 100$ entry at current accuracy is omitted: 0.712 was measured at $L = 30$, and the separately trained long-read run is not a substitute for it.

Per-base accuracy	L = 30	L = 100
0.712 (current, L=30 measured)	3.7×10^{-5}	
0.99	74%	37%
0.999	97%	90%

Quality scores and mappability. Per-base accuracy converts to a Phred quality by $Q = -10\log_{10}(1-\text{acc})$, so the 71.2% reported here is Q5.4, against a production floor of Q20 (99%) and a Q30 (99.9%) target for consensus over coverage. The gap is a viability gate rather than a quality metric, because a read is only useful if it maps, and the probability that a read is error-free falls as acc^L (Fig. 11).

At 71.2% and L = 30 a read carries approximately 8.6 errors, leaving error-free stretches of roughly three bases and a 3.7×10^{-5} chance of an error-free read (Table 2). Seed-and-extend mapping against a 3.2 Gb reference requires seeds long enough to be near-unique, and a three-base match occurs on the order of 10^7 times in a genome that size, so the read cannot be placed. The results indicate where the remaining accuracy sits. Temporal context is the largest demonstrated lever, since the same architecture reaches 89.6% per-base at L = 100, though realizing it requires read lengths the sensing geometry does not currently permit. The confusion structure points to a second, a fourth discrimination axis for the C/T pair, which accounts for the largest error channel and thymine's 45.0%. Consensus over coverage is a third in principle, but is unquantified here and bounded by the correlated residual below. The 71.2% figure is measured at L = 30 and is not extrapolated to any other length.

Caveats. Four qualifications bound these numbers. Single-call, not consensus: every figure is a per-event or per-base accuracy, and because the residual is correlated, consensus will not follow the naive binomial that would otherwise place it well above the per-base figure. Simulated ground truth: the accuracy characterizes the model and simulator jointly rather than a fabricated device, and its value lies in establishing that the three-axis

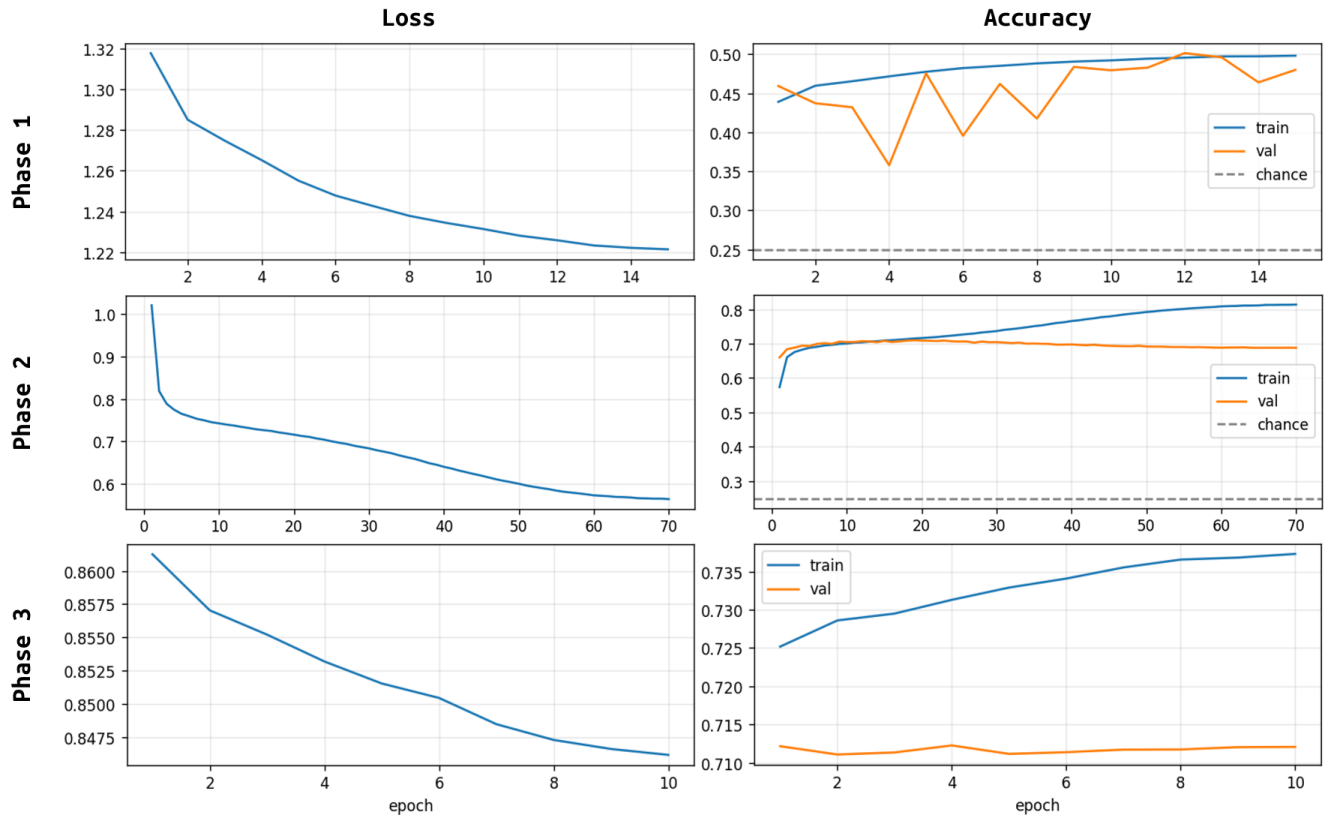

Fig. 9: Training curves by phase. Phase 1, encoder pre-training, 50.1% per-event. Phase 2, temporal stage over a frozen encoder, 71.0% per-base. Phase 3, end-to-end fine-tuning, flat at 71.2%.

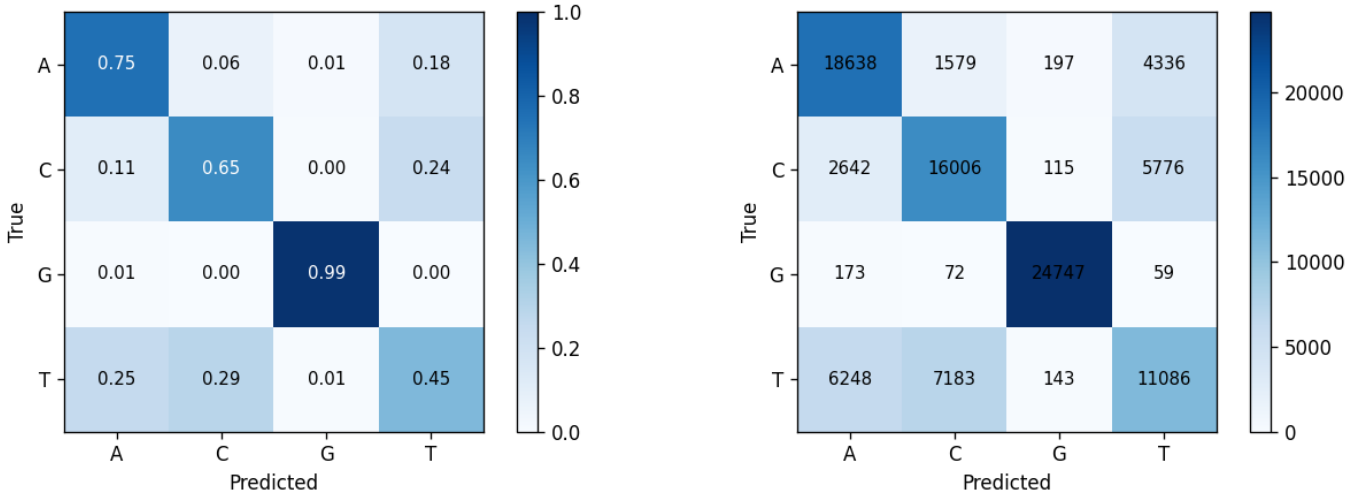

Fig. 10: Per-base confusion after fine-tuning, row-normalized (left) and as raw counts (right). Guanine resolves cleanly at 0.99 and the A↔G pair with it; the residual concentrates on the C↔T channel, with thymine called correctly in 0.45 of cases.

signal carries sufficient information to discriminate four bases, and in locating the residual where the physics predicts. Ensemble in phase by construction: the simulator implements the attrition dephasing bound rather than the phase-spread model, so these results do not test basecalling from a smeared ensemble in which

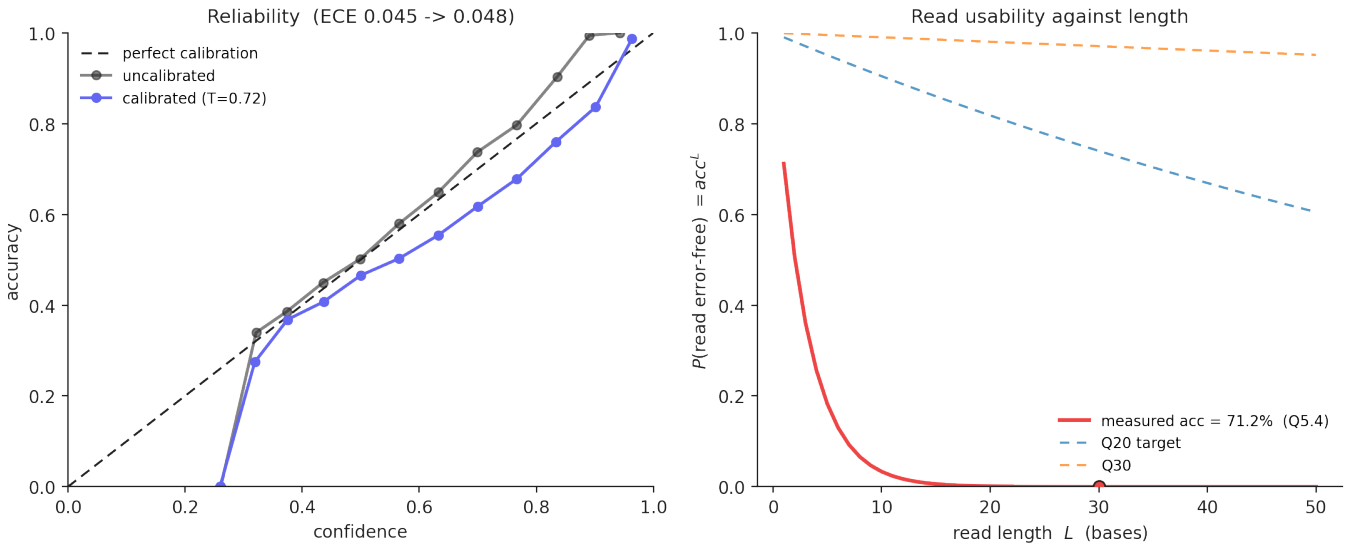

Fig. 11: Left: reliability of the raw posteriors, already calibrated at ECE 0.045, with temperature scaling making it marginally worse. Right: the viability gate. acc^L against read length at the measured accuracy and at the two Phred targets; the measured curve is at the floor well before the mapping length.

off-position copies contribute inter-symbol interference. Perfect segmentation: each figure assumes one spectrum per incorporation, a condition hardware does not provide, so the reported residual comprises substitutions alone, with insertions and deletions unrepresented.

Context ceiling. The decoder is not signal-limited so much as context-limited, and context is bounded not by the model but by the interval over which an incorporation event remains inside the electrode's screening envelope. Any mechanism that holds the growing strand within that envelope for longer therefore acts directly on accuracy, without modification to the sensor, the chemistry, or the decoder. The 89.6% obtained at L = 100 is what the present architecture already achieves given the context an extended envelope would supply. Read length and accuracy are consequently a single problem rather than two.

B. Long-Read Run ($L = 100$)

The numbers. L=100 exceeds the array's Debye sensing horizon and is not a supported operating point. What it isolates is decoder capacity given abundant context. The temporal stage reaches 89.6% per-base and the per-event encoder 76.7%, so the simulated signal carries enough information for a decoder to approach 90% given long context, which the physics does not allow with freestanding anchored DNA strands.

The faradaic axis, run over run. The third axis was not ablated under control at this read length. The matched faradaic-off arm was stopped three epochs into encoder pre-training, once its validation curve confirmed parity with the preceding dual-branch run, so what follows compares two separately trained runs rather than two arms of one. Read that way, adding the faradaic branch narrows the purine wall from roughly 12–13% to roughly 9% of A and G calls and lifts per-base accuracy from 88.2% to 89.6%; the per-event encoder moves further, from ~54% to 76.7%, because the temporal stage was already recovering most of the ambiguity the encoder leaves. The pyrimidine wall is unmoved, which is the same C/T residual the short-read run localizes. A completed within-run ablation, at either read length, remains outstanding.

The limit. The incorporation signal is observable only while the polymerase active site sits inside the electrode's screening envelope, roughly 3 nm or some nine bases at the ionic strength an active polymerase requires. Each base carries the active site 0.34 nm farther out, so the event leaves the sensing volume within tens of bases. This is interfacial electrostatics, not a limit of the model, and no amount of compute lifts it. The long-read number is therefore not a target to train toward. It is a measurement of what the decoder would give back if the sensing envelope were extended, and extending the envelope is an interfacial-engineering problem, worked through in Ap-

pendix F.2.

C. Data and Code Availability

The physics simulator that generates every dataset used here, the dataset definitions, and the basecaller with its training scripts are maintained as a single repository and are available from the authors on request. For a platform whose thesis is that the capability lives in the decoder rather than in the chip, the decoder and the model of the physics it inverts are the artifacts worth auditing, and the accuracies above are reproducible only against them.

VII. Discussion, Limitations, and Future Work

ECSEQ-0 is a passive electrode array read out by a learned decoder, developed here against its hardest application, label-free and wash-free sequencing by synthesis. What the design removes is most of what makes a sequencer expensive. There is no fluorescent label, no optical path, no camera, no flow cell consumed per run, and no reagent exchange between bases. What remains is a lithographically patterned surface held at a single temperature, a bulk reagent pool, and a model.

Two claims follow. The architectural claim is that reading incorporation electrically rather than optically moves the throughput bottleneck off the fluidics and onto readout electronics and compute. The figure of merit becomes an areal base rate independent of architecture, under which a single-run genome is a condition on parallelism rather than on time, contingent on an array scale not yet built. The physical claim is that the incorporation signal separates into three frequency-distinct axes carrying enough information to discriminate all four bases, with a faradaic axis introduced to break the purine pair. Both concern the array and its signal. The decoder and the application are the parts meant to change, which is the property that matters most. The instrument improves with the model, and the model improves without refabrication.

A. Limitations

Simulated data. All results are on simulated reads. The accuracy figures reflect the fidelity of the physics simulator rather than a measurement on hardware, and the transition to a real chip is a calibration-and-validation step whose outcome is not yet known.

The faradaic axis. The axis the model relies on for A↔G separation is the least established of the three. Its band assignment and the compatibility of elevated Au bias with polymerase fidelity are both modelling assumptions to be tested, and its contribution to accuracy rests on a comparison across two separately trained runs rather than on a completed control arm. Beyond either, the axis is destructive by mechanism, since guanine oxidation is irreversible and the measurement consumes the base it reads. That imposes a cumulative-exposure budget over a run and dilutes signal against previously oxidized bases still inside the sensing envelope, neither of which the reported accuracies model.

Read length. The incorporation signal is observable only inside the electrode's screening envelope, tens of bases, so ECSEQ-0 is a short-read architecture. Dephasing is a second ceiling at the same order, and on nominal kinetics the tighter of the two, which a measured per-copy cycle-time distribution will settle. The question it raises is not whether the cluster stays in phase, because it does not, but whether the basecaller can deconvolve a continuously smeared ensemble, which is nanopore basecalling in kind and has a known model class to attack it.

Surface reuse. The reuse estimate underpinning the cost argument is a mechanism rather than a measurement, and the multidentate anchor proposed to extend it is untested at operating conditions.

Correlated residual. The basecaller's error is structured rather than random. It concentrates on the C/T pair, thymine in particular, and traces to a single missing discrimination axis rather than to noise. The consequence is that coverage does not remove it, since the same error recurs at the same reference position across independent reads. Because the cause is absent information rather than excess noise, only added information resolves it, and a fourth discrimination axis for the pyrimidine pair is the decisive route. Calibrated posteriors consumed by a caller that models the confusion structure, and both-strand observation, reduce the impact without removing the cause.

Short reads. What a read must be is mappable, not long. Unique placement against a 3.2 Gb reference requires roughly 20 to 25 bases at sufficient per-base accuracy, well short of the 2×150 bp of established short-read platforms, and reference-based resequencing at that length

is well characterized rather than novel. Targeted panels, the platform's first buildable application, need less still, since the reference is a defined set of loci rather than a whole genome. De novo assembly and the spanning of long repeats are out of scope unless the sensing mechanism changes, and we do not claim them. Keeping this separate from the throughput analysis matters, because a read-length figure otherwise carries two unrelated claims, how fast a genome can be produced and what can be resolved once it is.

B. Read-Length, Accuracy, and Parallelism

Read length, per-base accuracy, and array parallelism are one constraint rather than three. Physics bounds read length, $L \leq \min(L_{\text{geom}}, L_{\text{coh}})$, with both terms landing at tens of bases. Accuracy then becomes a viability gate rather than a quality metric, since the probability a read is error-free is acc^L, which reaches 3.7×10^{-5} at 71.2% and L = 30, or roughly 8.6 errors per read. Throughput, $\eta P L \geq B$, is the one inequality short reads can still satisfy, and the only place a short read is compensable by pixel count.

ECSEQ-0 is therefore one architecture read at different points on a continuum. The targeted-panel device and the whole-genome sequencer are the same sensor, and what the instrument can address expands up that ladder as the physics matures and the array grows. Every rung is gated on the two-component mapping floor of length and accuracy, and every rung stays short-read, since whole-genome is reached by coverage rather than by long reads.

C. Beyond Sequencing

The same passive array and learned decoder define a broader instrument, and its most distinctive uses are the ones sequencing's read-length limit does not touch. The screening envelope bounds sequencing because synthesis carries the event away from the electrode. Something bound at the surface, a hybridized target, an antibody-antigen pair, a redox-active metabolite, sits inside that envelope permanently and is not read-length-limited at all. Because the readout is continuous and the bottleneck is compute rather than fluidics, such measurements can run as long-duration monitoring rather than single-shot tests, and each inherits the same compute-scaling advantage as the sequencer. Label-free pathogen detection, targeted molecular panels, and continuous multi-analyte biomarker monitoring are the natural next chips on the same reader, named here as directions the architecture supports rather than as results. What they share with sequencing is the entire stack beneath the application, the array, the reader, the decoder, and the simulator-and-calibration flywheel each new chip inherits from the last.

D. Next Experiment

The whole program gates on one measurement. Does a single natural-base incorporation, in a clonal cluster on a roughly 1 μm^2 Au/Pt pixel, produce a reproducible, base-discriminating impedance signature above the noise floor? Pre-registered gates convert that into a decision, with thresholds set from the blank-chip baseline before the run, since a threshold chosen after seeing the data is not a kill criterion.

Five risks are ranked. Cluster dephasing is gated on phase coherence to a pre-set cycle count on a known template, and failure sends the design to controlled-cycle operation, phase tracking, or a single-molecule fallback. Per-event signal-to-noise at 1 μm^2 is gated on mean $\Delta Z/Z$ clearing measured noise by a pre-set margin, and failure requires sensor redesign before scale-up. The molecular axis is gated on resolvable per-base dielectric contrast in the 10 to 100 kHz band, and failure points to a fourth discrimination axis. Sweep timing is gated on acquisition fitting inside the roughly 66 ms incorporation window, and failure calls for a pulsed or sub-banded schedule. Per-base accuracy is gated on reaching Q20 at the achievable read length, without which the architecture is not viable for resequencing.

The immediate work is hardware. Fabricate a single pixel, run the blank-chip and post-functionalization calibration, recalibrate the simulator against the measured interface, and fine-tune the basecaller onto real reads. Event segmentation is the next step on the model side, since the reported accuracies assume delimited events while the device delivers an unsegmented stream. Inferring the alignment rather than being given it is a CTC or transducer problem, the class that carries nanopore basecalling. Successive architectures on that unchanged physical signal moved read accuracy into the high eighties and low nineties, with the conditional-random-field decoder outperforming CTC [53,54].

If the per-event signal survives the move to hardware, the near-term instrument is a short-read electrochemical sensor with targeted DNA panels as its first application. The ladder above it is gated in order. First per-base accuracy to Q20, the threshold everything down-

stream rests on. Then array area, which decides whether a genome closes in one run. Then the multiplexed readout back end. Last sample preparation, which dominates wall-clock time but is borrowed and already met with margin. At the top is the aspiration that motivates the platform rather than a claim it makes today, massively parallel low-cost sequencing at population scale, the substrate a general genotype-to-phenotype model would ultimately require. The short-read instrument characterized here is the first step toward that array.

The array is the part that has to be built once. It is fixed, passive, and deliberately cheap, and its whole job is to report the state of its own surface honestly. Everything that turns that report into an answer is a model. An instrument built this way improves on the schedule of machine learning rather than that of semiconductor process development: the chip that reads a panel today reads it better next year, and reads something else entirely the year after, with no one returning to the fab in between. The chip is not the instrument. The model is, and the chip exists to give it something to look at.

References

[1] **Goodwin, S., J. D. McPherson, and W. R. McCombie (2016).** Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics* 17 (6): 333–351. DOI: 10.1038/nrg.2016.49 · PMID: 27184599.

[2] **Shendure, J., et al. (2017).** DNA sequencing at 40: past, present and future. *Nature* 550 (7676): 345–353. DOI: 10.1038/nature24286 · PMID: 29019985.

[3] **Wetterstrand, K. A.** DNA sequencing costs: data from the NHGRI Genome Sequencing Program (GSP). National Human Genome Research Institute. Last updated 16 May 2023. https://www.genome.gov/sequencingcostsdata Accessed 23 July 2026.

[4] **Zou, J., et al. (2019).** A primer on deep learning in genomics. *Nature Genetics* 51 (1): 12–18. DOI: 10.1038/s41588-018-0295-5 · PMID: 30478442.

[5] **Eraslan, G., et al. (2019).** Deep learning: new computational modelling techniques for genomics. *Nature Reviews Genetics* 20 (7): 389–403. DOI: 10.1038/s41576-019-0122-6 · PMID: 30971806.

[6] **Green, E. D., and M. S. Guyer (2011).** Charting a course for genomic medicine from base pairs to bedside. *Nature* 470 (7333): 204–213. DOI: 10.1038/nature09764 · PMID: 21307933.

[7] **Metzker, M. L. (2010).** Sequencing technologies: the next generation. *Nature Reviews Genetics* 11 (1): 31–46. DOI: 10.1038/nrg2626 · PMID: 19997069.

[8] **Rothberg, J. M., et al. (2011).** An integrated semiconductor device enabling non-optical genome sequencing. *Nature* 475 (7356): 348–352. DOI: 10.1038/nature10242 · PMID: 21776081.

[9] **Eid, J., et al. (2009).** Real-time DNA sequencing from single polymerase molecules. *Science* 323 (5910): 133–138. DOI: 10.1126/science.1162986 · PMID: 19023044.

[10] **Johnson, K. A. (2010).** The kinetic and chemical mechanism of high-fidelity DNA polymerases. *Biochimica et Biophysica Acta* 1804 (5): 1041–1048. DOI: 10.1016/j.bbapap.2010.01.006 · PMID: 20079883.

[11] **Fuller, C. W., et al. (2022).** Molecular electronics sensors on a scalable semiconductor chip: a platform for single-molecule measurement of binding kinetics and enzyme activity. *PNAS* 119 (5): e2112812119. https://doi.org/10.1073/pnas.2112812119 PMID: 35074874. Correction: *PNAS* 119 (29): e2205611119, https://doi.org/10.1073/pnas.2205611119.

[12] **Huang, S., et al. (2010).** Identifying single bases in a DNA oligomer with electron tunnelling. *Nature Nanotechnology* 5 (12): 868–873. https://doi.org/10.1038/nnano.2010.213 PMID: 21076404.

[13] **Tsutsui, M., M. Taniguchi, K. Yokota, and T. Kawai (2010).** Identifying single nucleotides by tunnelling current. *Nature Nanotechnology* 5 (4): 286–290. https://doi.org/10.1038/nnano.2010.42 PMID: 20305643.

[14] **Dontschuk, N., et al. (2015).** A graphene field-effect transistor as a molecule-specific probe of DNA nucleobases. *Nature Communications* 6: 6563. DOI: 10.1038/ncomms7563 · PMID: 25800494.

[15] **Flusberg, B. A., et al. (2010).** Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nature Methods* 7 (6): 461–465. DOI: 10.1038/nmeth.1459 · PMID: 20453866.

[16] **Daniels, J. S., and N. Pourmand (2007).** Label-free impedance biosensors: opportunities and challenges. *Electroanalysis* 19 (12): 1239–1257. https://doi.org/10.1002/elan.200603855 PMID: 18176631.

[17] **Long, Y.-T., et al. (2004).** Electrochemical detection of single-nucleotide mismatches: application of M-DNA. *Analytical Chemistry* 76 (14): 4059–4065. DOI: 10.1021/ac049482d · PMID: 15253643.

[18] **Ngavouka, M. D. N., et al. (2016).** Mismatch detection in DNA monolayers by atomic force microscopy and electrochemical impedance spectroscopy. *Beilstein Journal of Nanotechnology* 7: 220–227. DOI: 10.3762/bjnano.7.20 · PMID: 26977379.

[19] **Barsoukov, E., and J. R. Macdonald, eds. (2005).** *Impedance Spectroscopy: Theory, Experiment, and Applications.* 2nd ed. Hoboken, NJ: John Wiley & Sons.

[20] **Bard, A. J., and L. R. Faulkner (2001).** *Electrochemical Methods: Fundamentals and Applications.* 2nd ed. New York: John Wiley & Sons.

[21] **Love, J. C., et al. (2005).** Self-assembled monolayers of thiolates on metals as a form of nanotechnology. *Chemical Reviews* 105 (4): 1103–1170. DOI: 10.1021/cr0300789 · PMID: 15826011.

[22] **Herne, T. M., and M. J. Tarlov (1997).** Characterization of DNA probes immobilized on gold surfaces. *Journal of the American Chemical Society* 119 (38): 8916–8920. DOI: 10.1021/ja9719586.

[23] **Li, Z., et al. (2002).** Multiple thiol-anchor capped DNA-gold nanoparticle conjugates. *Nucleic Acids Research* 30 (7): 1558–1562. DOI: 10.1093/nar/30.7.1558 · PMID: 11917016.

[24] **Goode, J. A., J. V. H. Rushworth, and P. A. Millner (2015).** Biosensor regeneration: a review of common techniques and outcomes. *Langmuir* 31 (23): 6267–6276. DOI: 10.1021/la503533g · PMID: 25402969.

[25] **Shen, M. R., et al. (Illumina, Inc.) (2013).** Kinetic exclusion amplification of nucleic acid libraries. Patent WO2013188582A1 / US 8,895,249 B2.

[26] **Adey, A., et al. (2010).** Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. *Genome Biology* 11 (12): R119. DOI: 10.1186/gb-2010-11-12-r119 · PMID: 21143862.

[27] **Oxford Nanopore Technologies.** Rapid Sequencing gDNA (SQK-RAD004) protocol. nanoporetech.com/document/rapid-sequencing-sqk-rad004.

[28] **Collins, D. J., et al. (2015).** The Poisson distribution and beyond: methods for microfluidic droplet production and single cell encapsulation. *Lab on a Chip* 15 (17): 3439–3459. DOI: 10.1039/c5lc00614g · PMID: 26226550.

[29] **Drmanac, R., et al. (2010).** Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science* 327 (5961): 78–81. DOI: 10.1126/science.1181498 · PMID: 19892942.

[30] **Margulies, M., et al. (2005).** Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437 (7057): 376–380. DOI: 10.1038/nature03959 · PMID: 16056220.

[31] **Adessi, C., et al. (2000).** Solid phase DNA amplification: characterisation of primer attachment and amplification mechanisms. *Nucleic Acids Research* 28 (20): e87. DOI: 10.1093/nar/28.20.e87 · PMID: 11024189.

[32] **Johnson, R. C., et al. (1996).** Electron-correlated calculations of electric properties of nucleic acid bases. *Journal of Physical Chemistry* 100 (48): 18875–18881. DOI: 10.1021/jp962161o.

[33] **Keighley, S. D., et al. (2008).** Optimization of DNA immobilization on gold electrodes for label-free detection by electrochemical impedance spectroscopy. *Biosensors and Bioelectronics* 23 (8): 1291–1297. DOI: 10.1016/j.bios.2007.11.012 · PMID: 18178423.

[34] **Laborde, C., F. Pittino, H. A. Verhoeven, S. G. Lemay, L. Selmi, M. A. Jongsma, and F. P. Widdershoven (2015).** Real-time imaging of microparticles and living cells with CMOS nanocapacitor arrays. *Nature Nanotechnology* 10 (9): 791–795. https://doi.org/10.1038/nnano.2015.163.

[35] **Kulkarni, G. S., and Z. Zhong (2012).** Detection beyond the Debye screening length in a high-frequency nanoelectronic biosensor. *Nano Letters* 12 (2): 719–723. https://doi.org/10.1021/nl203666a.

[36] **Steenken, S., and S. V. Jovanovic (1997).** How easily oxidizable is DNA? One-electron reduction potentials of adenosine and guanosine radicals in aqueous solution. *Journal of the American Chemical Society* 119 (3): 617–618. DOI: 10.1021/ja962255b.

[37] **Oliveira-Brett, A. M., V. Diculescu, and J. A. P. Piedade (2002).** Electrochemical oxidation mechanism of guanine and adenine using a glassy carbon microelectrode. *Bioelectrochemistry* 55 (1–2): 61–62. DOI: 10.1016/S1567-5394(01)00147-5 · PMID: 11786341.

[38] **Vogiazi, V., et al. (2021).** Effects of experimental conditions on the signaling fidelity of impedance-based nucleic acid sensors. *Analytical Chemistry* 93 (2): 812–819. DOI: 10.1021/acs.analchem.0c03269.

[39] **Arslan, S., et al. (2024).** Sequencing by avidity enables high accuracy with low reagent consumption. *Nature Biotechnology* 42 (1): 132–138. DOI: 10.1038/s41587-023-01750-7 · PMID: 37231263.

[40] **Deamer, D., M. Akeson, and D. Branton (2016).** Three decades of nanopore sequencing. *Nature Biotechnology* 34 (5): 518–524. DOI: 10.1038/nbt.3423 · PMID: 27153285.

[41] **Jain, M., et al. (2016).** The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biology* 17: 239. DOI: 10.1186/s13059-016-1103-0 · PMID: 27887629.

[42] **Loman, N. J., et al. (2012).** Performance comparison of benchtop high-throughput sequencing platforms. *Nature Biotechnology* 30 (5): 434–439. DOI: 10.1038/nbt.2198 · PMID: 22522955.

[43] **Quail, M. A., et al. (2012).** A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* 13: 341. DOI: 10.1186/1471-2164-13-341 · PMID: 22827831.

[44] **Stoddart, D., et al. (2009).** Single-nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore. *PNAS* 106 (19): 7702–7707. DOI: 10.1073/pnas.0901054106 · PMID: 19380741.

[45] **Rang, F. J., W. P. Kloosterman, and J. de Ridder (2018).** From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy. *Genome Biology* 19: 90. DOI: 10.1186/s13059-018-1462-9 · PMID: 30005597.

[46] **Kokoris, M., R. McRuer, M. Nabavi, et al. (2025).** Sequencing by Expansion (SBX) — a novel, high-throughput single-molecule sequencing technology. *bioRxiv*, v2 posted 2025-03-15 (preprint, not peer-reviewed). DOI: 10.1101/2025.02.19.639056.

[47] **Erlich, Y., et al. (2008).** Alta-Cyclic: a self-optimizing base caller for next-generation sequencing. *Nature Methods* 5 (8): 679–682. DOI: 10.1038/nmeth.1230 · PMID: 18604217.

[48] **Kircher, M., U. Stenzel, and J. Kelso (2009).** Improved base calling for the Illumina Genome Analyzer using machine learning strategies. *Genome Biology* 10 (8): R83. DOI: 10.1186/gb-2009-10-8-r83 · PMID: 19682367.

[49] **Whiteford, N., et al. (2009).** Swift: primary data analysis for the Illumina Solexa sequencing platform. *Bioinformatics* 25 (17): 2194–2199. DOI: 10.1093/bioinformatics/btp383 · PMID: 19549630.

[50] **Ledergerber, C., and C. Dessimoz (2011).** Base-calling for next-generation sequencing platforms. *Briefings in Bioinformatics* 12 (5): 489–497. DOI: 10.1093/bib/bbq077 · PMID: 21245079.

[51] **Louizos, C., M. Welling, and D. P. Kingma (2018).** Learning sparse neural networks through L0 regularization. *ICLR 2018.* arXiv: 1712.01312

[52] **Balın, M. F., A. Abid, and J. Zou (2019).** Concrete autoencoders: differentiable feature selection and reconstruction. *ICML 2019*, PMLR 97: 444–453.

[53] **Wick, R. R., L. M. Judd, and K. E. Holt (2019).** Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biology* 20: 129. https://doi.org/10.1186/s13059-019-1727-y · PMID: 31234903

[54] **Pagès-Gallego, M., and J. de Ridder (2023).** Comprehensive benchmark and architectural analysis of deep learning models for nanopore sequencing basecalling. *Genome Biology* 24: 71. https://doi.org/10.1186/s13059-023-02903-2.

[55] **Kesler, V., B. Murmann, and H. T. Soh (2020).** Going beyond the Debye length: overcoming charge screening limitations in next-generation bioelectronic sensors. *ACS Nano* 14 (12): 16194–16201. https://doi.org/10.1021/acsnano.0c08622.

[56] **Piccinini, E., et al. (2018).** Pushing the boundaries of interfacial sensitivity in graphene FET sensors: polyelectrolyte multilayers strongly increase the Debye screening length. *Journal of Physical Chemistry C* 122 (18): 10181–10188. https://doi.org/10.1021/acs.jpcc.7b11128.

[57] **Gao, N., W. Zhou, X. Jiang, G. Hong, T.-M. Fu, and C. M. Lieber (2015).** General strategy for biodetection in high ionic strength solutions using transistor-based nanoelectronic sensors. *Nano Letters* 15 (3): 2143–2148. https://doi.org/10.1021/acs.nanolett.5b00133.

[58] **Gao, N., et al. (2016).** Specific detection of biomolecules in physiological solutions using graphene transistor biosensors. *PNAS* 113 (51): 14633–14638. https://doi.org/10.1073/pnas.1625010114.

[59] **Shoorideh, K., and C. O. Chui (2014).** On the origin of enhanced sensitivity in nanoscale FET-based biosensors. *PNAS* 111 (14): 5111–5116. https://doi.org/10.1073/pnas.1315485111.

[60] **Olsen, T. J., Y. Choi, P. C. Sims, O. T. Gul, B. L. Corso, C. Dong, W. A. Brown, P. G. Collins, and G. A. Weiss (2013).** Electronic measurements of single-molecule processing by DNA polymerase I (Klenow fragment). *JACS* 135 (21): 7855–7860. https://doi.org/10.1021/ja311603r.

Appendix A. Contributions

The authors contributed to the following areas.

Sensor architecture and fabrication. Michael Vaden, Evan Goldstein

Surface chemistry and functionalization. Jack Ready

Signal model and physics simulator. Michael Vaden

Basecaller design and training. Michael Vaden

Readout scaling, throughput, and cost analysis. Evan Goldstein, Michael Vaden

Appendix B. Randles–CPE Systems

Electrodes in contact with a buffer form an electrical double layer, which is a compact Helmholtz layer and a diffuse Gouy-Chapman layer of counter-ions. To an AC stimulus the interface behaves as a double-layer capacitance in parallel with a charge transfer pathway. In series with the solution resistance this yields the Randles model.

Because real electrodes are rough and have a distribution of relaxation times, the ideal capacitor is replaced by a constant-phase element.

$$Z_{\text{layer}}(\omega) = R_s + \frac{1}{Q\,(j\omega)^{\alpha} + \dfrac{1}{R_{\text{ct}} + Z_W(\omega)}} \qquad (2)$$

Each element is physical: R_s is the bulk electrolyte path; Q and α describe the double-layer charging; R_{ct} gates faradaic electron transfer across the interface; and Z_W (see Warburg diffusion) captures the diffusion-limited tail. An incorporation event moves one or more of these elements; EIS resolves the move frequency-by-frequency.

Appendix C. Local Incorporation

A correct dNTP incorporation is not a single instantaneous step but a sequence of physical changes localized at the SAM-tethered primer lawn, each of which touches the double layer. Phosphodiester bond formation releases a pyrophosphate (PPi) anion and a proton, both charged species appearing within nanometres of the electrode, transiently shifting local ionic strength and the charge balance of the double layer. The screening cloud of counter-ions then re-equilibrates around the new charge state of the primer terminus, modulating C_{dl}. Meanwhile the polymerase cycles through nucleotide recognition in the open state and closed-fingers catalysis, the two together making up the dwell, and the incorporated nucleobase changes the steric and dielectric environment at the thiol-SAM surface.

Three circuit elements move. The double-layer CPE (Q) is the dominant observable: ionic rearrangement and the conformational change directly modulate double-layer charging, and at the 1 μm^2 pixel scale area scaling pushes R_{ct} very high, so the parallel combination collapses onto the CPE branch and the measurable ΔZ/Z is carried almost entirely by ΔQ. The charge-transfer resistance (R_{ct}) is faradaic, moving when the interface actually passes electrons, as in base-intrinsic oxidation under bias; it is negligible at benign potential but the deliberate handle behind the faradaic discrimination axis detailed below. The Warburg term (Z_W) moves with the transient diffusion of the released ionic species, dominates the low-frequency tail, and blends into the CPE as $\alpha \to 0.5$.

A DC measurement collapses the interface to a single number and cannot separate mechanisms that live at different timescales, whereas AC impedance spectroscopy resolves the response across frequency, where the physical mechanisms map cleanly onto bands. The lower bound of a usable sweep is set by the event window: the fastest base's dwell (~66 ms, adenine) means a frequency is only measurable if the event outlasts ~1/f, fixing a practical floor near 100 Hz. Frequency resolution is therefore not a convenience; it is what makes the distinct physical mechanisms separable at all.

Appendix D. Faradaic Discrimination Axis

Motivation. The Pt kinetic axis reflects polymerase conformation, and A and G share overlapping dwell distributions; the Au molecular axis as currently modelled is a non-faradaic dielectric/dipole effect and does not, on its own, encode the feature that most strongly separates A from G. The faradaic axis supplies that feature, and in simulation it does its job: guanine is called at 98.8% accuracy and the A↔G purine pair is essentially resolved.

Physical basis. The natural bases have distinct intrinsic oxidation potentials: guanine is most readily oxidized (~+0.9 to +1.0 V vs Ag/AgCl), adenine higher (~+1.1 to +1.2 V), and cytosine and thymine both oxidatively sluggish at higher, closely spaced potentials. Biasing the

Au charge-transfer electrode toward the guanine oxidation onset makes the faradaic charge-transfer resistance R_{ct} base-dependent: guanine presents a smaller charge-transfer arc than adenine. This is a base-intrinsic electron-transfer signal, separate from the dielectric/dipole signal already attributed to the Au layer.

Implementation. The Au layer is already the charge-transfer working electrode and the DNA anchor, so no new electrode is required; the change is how it is biased and read. Hold or pulse the Au interface toward the guanine oxidation onset and resolve the low-frequency charge-transfer arc as a feature family separate from the capacitive/dielectric response. In the CPE-Randles description above this is a base-dependent R_{ct} element controlling the low-frequency arc.

Frequency floor. The charge-transfer arc resolves at low frequency, but the per-event dwell bounds how low the sweep can reach. At the ~66 ms shortest dwell (adenine) the floor is ~15–30 Hz, so only the high-frequency shoulder of the arc is accessible within a single incorporation, not the full sub-Hz R_{ct} plateau, so the accessible faradaic signal is partial.

Operating point. Biasing toward guanine oxidation is elevated-potential relative to the benign double-layer measurement; the cost to polymerase fidelity and DNA integrity is unquantified and must be measured on the bench. A short faradaic interrogation pulse, separate from the continuous low-bias sweep, is one way to limit enzyme exposure and should be compared against continuous biasing.

Mediator-free. A solution redox mediator (for example a Ru complex shuttling guanine oxidation) would sharpen the signal but reintroduces an added reagent and breaks the label-free thesis. This axis is mediator-free by design.

Appendix E. Scaling Readout

The throughput gains EIS sequencing can provide require a corresponding parallelization of the readout, and are bounded by it. At a 2 µm pitch, panel scale is a 900 mm^2 active area holding 2.25×10^8 pixels, and every figure in this appendix is quoted there. That array does not close a genome in one pass at a supported read length; §E.2 works the parallelism condition $\eta PL \geq B$ through and finds that a single run needs a larger array or a finer pitch. What the panel-scale numbers establish is where the binding constraint sits, and it is not the array. Pixel count is not the constraint on throughput; reading that many pixels out fast enough is.

E.1 Channel Count and Multiplexing

To call a base, a pixel's chemistry-window spectrum must be captured during the incorporation window. The shortest window is adenine at ~66 ms. One multisine acquisition at the 100 Hz measurement floor takes ~10 ms (a 100 Hz component cannot be measured faster than its own 10 ms period). Reading all 225 M pixels at the ~15 Hz frame rate (the inverse of adenine's 66 ms incorporation window) implies 3.4×10^9 spectra/s (1 spectrum/pixel/frame $\times$ 225 M pixels $\times$ 15 frames/s). A single time-shared front-end delivers on the order of 100 spectra/s, which puts the naive requirement at ~34 million independent front-ends. That is not manufacturable at a cost consistent with a passive array.

Multiplexing breaks the one-pixel-per-front-end assumption. Drive many pixels simultaneously with orthogonal excitations (distinct carrier frequencies, or Hadamard codes across rows), sum the responses onto a shared sense line, and demultiplex digitally after one fast ADC. This is standard practice in SQUID/MRI arrays, radio astronomy, and large capacitive sensors. One ADC then carries M pixels concurrently, where $M \approx$ ADC bandwidth / per-pixel bandwidth.

Let $S = N_{\text{ADC}}M$ be the number of concurrent measurement slots. Serving the whole array at cadence requires $S = 3.4 \times 10^9/100 = 3.4 \times 10^7$ slots, which Table 3 works through scheme by scheme.

~6,800 fast ADCs is an ordinary data-acquisition back end, comparable in channel count to a single high-density nanopore flow cell rather than to an aggregate multi-flow-cell installation. The SNR-safe corner below, at $M \approx 512$, costs roughly 67,000 converters for the same coverage; that is still buildable and is the operating point to quote.

E.2 Metrics Modelling

The answer turns on one observation. Read length does not appear in the base rate at all. It enters only through the run count, through how many times R the fixed per-run overhead T_{prep} must be paid. A short read is therefore compensable by pixel count, and the only condition for a one-run genome is

$$\eta PL \geq B \tag{3}$$

Table 3: Multiplexing schemes against the concurrent-slot requirement, and whether each is buildable.

Scheme	ADCs	Mux depth	Slots	Buildable?
Naive (no mux)	34 M	1	34 M	no
FDM, 100 MSPS ADCs	~68,000	~500	34 M	yes, many boards
FDM, 1 GSPS ADCs	~6,800	~5,000	34 M	yes, comfortably

Table 4: Active area needed to close a 30× genome in one run, by read length and pixel pitch.

Read length L	@ 2 µm	@ 1 µm	@ 0.5 µm
10	104,000 mm^2	26,000 mm^2	6,500 mm^2
30	34,600 mm^2	8,650 mm^2	2,160 mm^2
100	10,400 mm^2	2,600 mm^2	650 mm^2

where η is the occupancy yield. This matters because randomly seeding discrete sites caps single-occupancy near 37% by Poisson statistics; using the un-derated P overstates capacity by 2.7×. We carry η symbolically, defaulted to 0.37, with active-bias loading control as the upside case rather than the assumption.

Consider the duty cycle, $L\,t_{\text{base}}/(T_{\text{prep}} + L\,t_{\text{base}})$: at the supported $L = 30$, sequencing occupies under 0.2% of wall-clock time, and even at a read an order of magnitude longer it stays under 3%. Sample preparation dominates by hundreds of times. Whenever $R = 1$, therefore, $T_{\text{genome}} \approx T_{\text{prep}}$, and the sub-hour genome is a *prep-time and array-area* claim rather than a read-length claim.

At 900 mm^2 the single-run genome does not close: $\eta PL \geq B$ would require $L \geq 1,150$. It requires a larger array or a finer pitch. The area that does close it at a supportable read length follows from the same inequality (Table 4), whose inputs are collected in Table 5.

At a supportable $L = 30$ and 0.5 µm pitch a one-run genome needs roughly 2,200 mm^2, about 2.4× the 900 mm^2 panel, and within the reticle-stitched area of a single 300 mm wafer. Under that configuration $T_{\text{genome}} = 35\,min + 30 \times 0.1175\,s \approx 35$ min. The sequencing term has become negligible: 3.5 seconds against a 35-minute prep.

$$T_{\text{genome}} = R\left(T_{\text{prep}} + L\,t_{\text{base}}\right)$$
$$R = \left\lceil \frac{B}{\min(P,\, d\,S)\,L} \right\rceil \tag{4}$$

The back end must read every active pixel each frame: with $S = N_{\text{ADC}}M$ concurrent slots and a time-sharing gain $d = 6.6$, covering all $P = 2.25 \times 10^8$ pixels requires $dS \geq P$. At the SNR-safe multiplex depth recommended below ($M \approx 512$) that is $\approx 6.7 \times 10^4$ converters; the throughput-maximal corner ($M = 5,000$, $N_{\text{ADC}} \approx 6,800$, $S = 3.4 \times 10^7$) reaches the same coverage with far fewer converters, but at a multiplex depth the SNR budget forbids.

E.3 SNR Under Multiplexing

Multiplexing does not create information, it trades the solved channel-count problem for an SNR problem. Stacking pixels onto one line and ADC shares the converter's dynamic range, since each pixel keeps $16 - \log_2 M$ effective bits of a 16-bit ADC ($M = 5,000 \Rightarrow$ ~4 bits), and crosstalk/intermodulation leaks carriers into each other. Two quantities must be distinguished. The ~15% impedance shift is the molecular-axis perturbation, the always-on dielectric signal every event carries; it is not the faradaic discriminant. Biasing toward the guanine oxidation onset collapses R_{ct} from its SAM-blocked baseline by two orders of magnitude, roughly 99%, and that collapse fires on every biased event regardless of base. What actually discriminates is the base-dependent *ratio* riding on top of it, about 1.67× between guanine and the pyrimidines. The converter must therefore span the full oxidation-onset swing, roughly $\log_2(100,000/555) \approx 7.5$ bits, before spending the further $\log_2(1.67) \approx 0.7$–1 bit that carries base identity.

Time-to-genome depends only on $S = N_{\text{ADC}} \times M$, while SNR headroom depends only on M, so the same speed can be bought two ways: deepen the mux (raise M, fast and cheap, but headroom falls as $16 - \log_2 M$), or add converters (raise N_{ADC} at fixed M, same speed at the same SNR, with BOM scaling ~linearly since each converter then needs only ~tens of MSPS). The plane also has a ceiling (Fig. 12), and it is the more useful feature. Once dS covers the array every further converter is idle, and the run count bottoms out at $\lceil B/(\eta PL)\rceil$, which at panel scale and $L = 30$ is 39 runs, about a day on one reusable chip. The three marked points all sit on that boundary and therefore all cost the same time; they differ only in SNR headroom and bill of materials. $M = 5,000$ reaches it with ~6,800 converters at roughly 4 effective bits per pixel, the recommended $M = 512$ with ~67,000 at 7,

Table 5: Inputs to the time-to-genome model.

Symbol	Meaning	Value
B	bases needed (30× of 3.2 Gb)	96 Gb
P	pixels (panel scale)	2.25×10^8
t_{base}	per-base polymerase time	0.1175 s
S	concurrent measurement slots	design variable
T_{prep}	one-time sample prep + amplification	~35 minutes
t_{A}	shortest incorporation time (base A)	66 ms
t_{acq}	one acquisition at the 100 Hz floor	10 ms
d	pixels served per slot per frame (= 66 / 10; 3.3 on a pulsed faradaic schedule)	6.6

and an SNR-safe $M = 100$ with ~341,000 at 9.5. The sweet spot is $M \approx 100$–600. What the plane rules out is buying speed with readout: at this array size the genome is array-limited rather than readout-limited, and the lever that moves it is the area of §E.2, not the converter count.

A fourth risk sits underneath all of these and is specific to the faradaic axis. Because guanine oxidation is irreversible, every previously interrogated guanine still resident inside the Debye sensing envelope continues to present a low-R_{ct}, guanine-like contribution at every subsequent cycle. Within a 9–26 base envelope at GC-neutral composition the expected count of such residents is $n_G \approx 2$–6, and simulation at the working envelope gives a mean of 2.9. The marginal signal attributable to the newest incorporation is therefore diluted roughly as $1/(1+n_G)$, a factor of three to four, against a background of its own prior bases.

E.4 Passive-Matrix Interconnect Limits

A 900 mm^2 array at 2 µm pitch is roughly 30 mm on a side, so a single row or column line runs 30 mm and carries about 15,000 pixel taps. The array is passive by design, with no per-pixel buffer or amplifier, which is what makes the chip cheap and reusable, so every unselected pixel loads the shared sense node with its own parasitic capacitance, and the thin-film interconnect contributes distributed series resistance over the full run. The resulting RC network has a corner frequency that falls as the line lengthens, competing directly with the 100 kHz top of the main sweep, which is precisely where the molecular axis has its signal.

Two features separate this from the ADC-sharing risk. It scales with *array size* rather than mux depth, so the large, shallow-mux configurations that resolve the dynamic-range problem do not avoid it and may worsen it. And it is a property of the passive-array choice itself: an active-matrix design would buffer each pixel and largely remove it, at the cost of a transistor per site and the fabrication simplicity the platform depends on. We have not modelled line resistance, per-line capacitance, or the resulting bandwidth derating at any scale, and the prototype's 4-pixel direct-wired geometry cannot exhibit the effect. It is an unquantified risk.

E.5 Measurement Frequency as a Sensing-Depth Lever

Ionic screening is not instantaneous. Below roughly 1 MHz, ions have ample time to screen normally [55], and ECSEQ-0's present sweep tops out at 100 kHz, squarely inside the regime where the full screening penalty applies. Detection beyond the quasi-static screening length has been reported above 1 MHz [35], and a CMOS nanocapacitor array measuring per-pixel interfacial impedance has shown sensing depth extending measurably as sampling rises toward 50 MHz [34].

Two caveats bound this. The characteristic ionic relaxation rate scales as D/κ^{-2}, which for $D \approx 10^{-9}$ m^2/s and $\kappa^{-1} \approx 1$ nm places the true crossover in the hundreds of MHz, far above the 1 MHz results, whose mechanism is better described as nonlinear mixing with molecular dipoles than as outrunning the ion cloud. And those results have not been independently replicated. We therefore treat extension of the readout band toward the MHz decade as an identified and inexpensive lever to test on single-pixel hardware, not a solved path.

E.6 Data and Compute Budget

The central architectural claim is that removing the label and the wash moves the bottleneck from fluidics onto readout electronics and compute. Both figures below are generated from the model definition rather than asserted, and both are projections at array scales that have not been built.

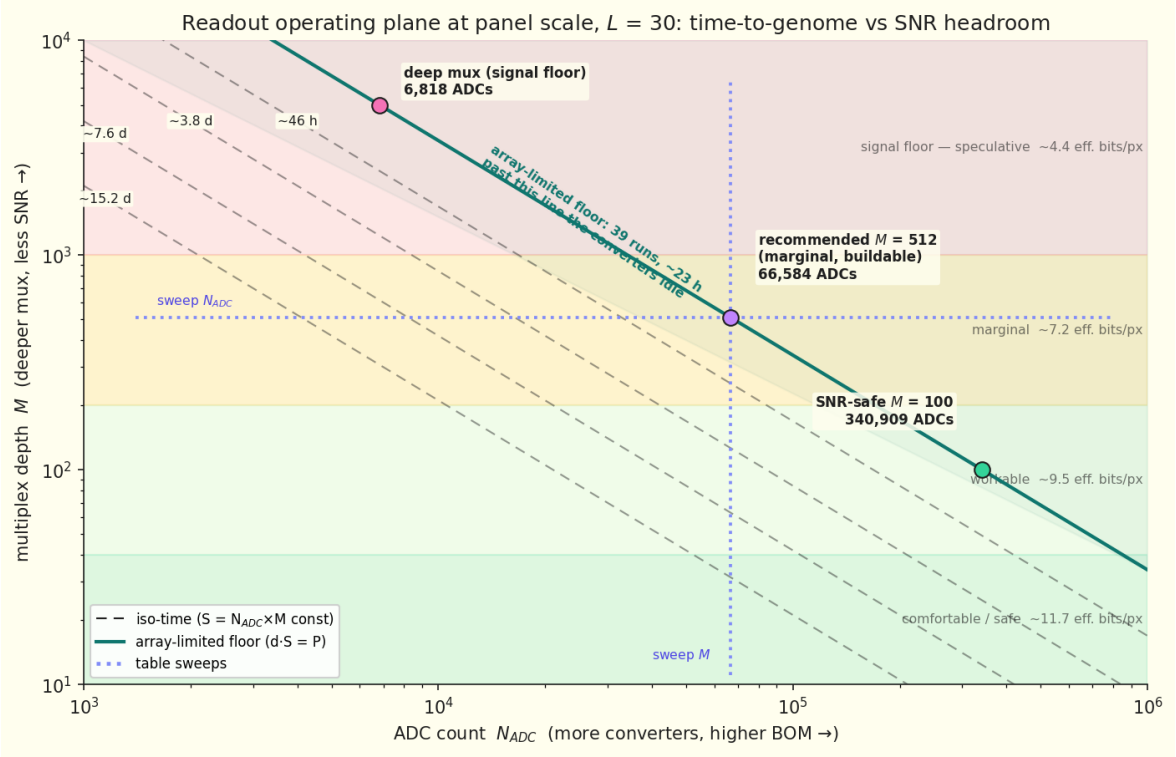

Fig. 12: The readout operating plane at panel scale and $L = 30$. Diagonals are equal time-to-genome, horizontal bands per-pixel SNR headroom, dotted lines the two sensitivity sweeps, and the solid boundary the array-limited floor past which the readout is no longer the constraint. All three marked points sit on it, so they differ in SNR and bill of materials but not in time.

Table 6: Base rate and basecalling compute at the three device scales.

Scale	Bases/s	Basecalling
ECSEQ-0 prototype (4 px)	34	1.5 GFLOP/s
ECSEQ-0 full-spec die	7.8×10^{6}	354 TFLOP/s
Panel scale	1.9×10^{9}	86 PFLOP/s

Base rate. The architecture-independent figure of merit is $\eta P/t_{\text{base}}$. At panel scale with $\eta = 1$ this is 1.9×10^9 bases/s, or about 2.1×10^6 bases/s/mm^2 of active area. Unlike read length, this survives every correction to read length: it depends only on pixel count and polymerase turnover.

Front-end data rate. Each event yields one 62-point complex spectrum, or 124 values. At panel scale and 16-bit resolution that is ~475 GB/s off the array; post-demodulation at 8 bits, ~237 GB/s. The reduction chain matters more than either number: raw converter samples are demodulated per tone at the front end, reduced to the 124-value perturbation vector, and only then leave the instrument. That reduction must happen in acquisition hardware, not on a host.

Basecalling compute. The per-event encoder costs 19.3 MMAC (38.7 MFLOP) per spectrum; the temporal stage adds 6.5 MFLOP per base at a 30-base read. Together, 5.4 M parameters. Table 6 carries both figures across the three device scales.

The full-spec demonstrator basecalls in real time on a single contemporary GPU server. Scaling the instrument means scaling compute, which sits on a cost-and-capability curve that fluidics does not.

E.7 Readout Feasibility: an End-to-End Simulation

Two questions remain open: whether a shared-converter chain preserves the basecall once every non-ideality is included, and whether an SNR-safe multiplex depth is reachable at all. A first-principles simulation study of the full readout chain addresses both. It models the electronics between the pixel and the basecaller as five stages: array substrate, excitation (drive), front-end (sense), code-division multiplex, and digital demux, each allocated a slice of an end-to-end error budget and graded in one currency (effective bits) against the only metric that matters:

does the recovered $(\Delta log|Z|,\ \Delta\varphi)$ vector still separate the bases. Every result is simulation rather than hardware, and the basecaller seam is stubbed, so accuracy is graded by a data-domain separability lower bound rather than the trained model.

The composed chain preserves the basecall. End to end, the recovered perturbation is 0.88 effective-bit against a 1.39-bit budget, degrading per-event separability (NCM) by 0.022 and consensus-at-30× by 0.009, with no disproportionate cross-group error. Every stage lands within its allocation, and one dominates: the analog front-end and quantization carry ~0.70 bit, about 80% of the composed degradation and an order of magnitude above any other stage. The binding constraint on the whole readout is therefore the front-end noise at the acquisition integration time, not multiplex depth, excitation, or demux.

Analog nulling. The dynamic-range penalty booked above ($16-\log_2 M$ bits) assumes the converter must span the full ~$10^{10}\,\Omega$ baseline. An analog baseline nuller, the inverted quiescent comb subtracted before digitization, standard in SQUID-FDM readouts, cancels that baseline and reclaims ~4.85 bits, so the ADC digitizes only the ~2.7-bit perturbation. With it, the buildable depth set by ADC dynamic range rises to ~7,100 pixels per converter; without it, it collapses to ~250. The nuller is therefore a precondition, not an optimization, and its binding real-world requirement is a recalibration cadence holding baseline drift to $\leq$0.52% between calibration and event.

Buildable operating point. The study's recommended configuration is $M \approx 512$ pixels per 16-bit converter under Hadamard code-division multiplexing, which lands inside the SNR-safe sweet spot ($M \approx$ 100–600) independently identified above, and well below the ~7,100 dynamic-range ceiling. At that depth the full 2.25×10^8-pixel array needs $\approx 6.7 \times 10^4$ converters, a ~3,400× reduction from one-per-pixel. The study itself quotes $\approx 4.4 \times 10^5$ for the same array at the same depth, because it books one concurrent slot per pixel; the figure here additionally applies the $d = 6.6$ time-sharing gain of §E.2, under which one slot acquires 6.6 pixels inside a single incorporation window. The two numbers differ only in whether time-sharing is counted, and the converter count is the one quantity in this appendix where that distinction changes the answer. Code-division is chosen over the frequency-division assumed earlier because the flat ±1 code envelope carries no peak-to-average penalty (frequency-division loses up to ~24× crest-limited SNR at this depth), spends the converter's full range on signal every frame, and demultiplexes exactly by a Walsh–Hadamard transform.

E.8 Cost Structure

Optical and most electronic platforms sequence on a single-use flow cell, chip, or SMRT cell: substrate, surface preparation, and fluidic consumable are paid in full every run. ECSEQ-0's thiol-Au primer lawn is regenerated between runs by a denaturing strip that leaves the anchor intact, so the chip is paid for once and amortized over its 50–100-run reuse life; the per-run consumable is then the borrowed short-read library prep, not the chip.

Two properties lower the amortized cost further. The chip is not sequencing-specific but a general-purpose EIS array, so its fabrication cost spreads across every application it serves, not sequencing runs alone: a chip that also runs molecular panels or continuous monitoring between sequencing runs amortizes over a far larger denominator. And it is passive and label-free, with no per-pixel transistors, optics, fluorophores, or wash hardware, so its bill of materials is a thin-film electrode array plus the off-chip reader, and the reader is reused across every chip.

To first order the per-genome cost is $C_{\text{chip}}/N_{\text{reuse}} + C_{\text{prep}}$, both terms small and neither a single-use flow cell.

Appendix F. Read-Length Limits: Interfacial Geometry and Cluster Coherence

Every figure in the throughput analysis rests on a read length L, and L is not a free parameter. Two independent mechanisms bound it.

$$L_{\max} = \min\left(L_{\text{geom}},\ L_{\text{coh}}\right)$$
$$L_{\text{geom}} \approx \frac{3\kappa^{-1}}{(0.34\,\text{nm/bp})\sin\theta} \tag{5}$$

L_{geom} is the interfacial sensing horizon: the polymerase active site must remain inside the electrode's sensing volume for a base to be discriminated at all, and that volume is bounded by the Debye screening length κ^{-1}. L_{coh} is the cluster-coherence horizon: the copies must stay in positional register for their summed signal to resolve one incorporation per position.

The two land at the same order, and neither is comfortably slack. Geometry gives roughly nine bases in a buffer that supports an active polymerase. The phase-spread account of coherence gives roughly four on nominal kinet-

Table 7: Debye screening length and the resulting sensing envelope, by buffer ionic strength.

Ionic strength	κ^{-1}	$3\kappa^{-1}$	Bases in envelope
90 mM (working)	1.0 nm	3.0 nm	~9
50 mM	1.4 nm	4.1 nm	~12
10 mM (dilute)	3.0 nm	9.1 nm	~26
1 mM	9.6 nm	28.8 nm	~85 (no polymerase activity)

ics, so on the numbers derived below coherence is the tighter of the two, and lifting either one alone buys nothing. Which of them actually binds turns on the per-copy cycle-time distribution, which has not been measured. The subsections below derive the geometric horizon, survey what is known about moving it, and then derive the coherence horizon. Both remain unmeasured; both are single-pixel validation targets.

F.1 Geometric Sensing Horizon

The sensing thesis of this paper is that the incorporation event perturbs the interfacial impedance. That perturbation is only observable while the polymerase active site sits inside the region where the electrode's field is not yet screened by mobile ions. For a symmetric electrolyte, $\kappa^{-1} = 0.304/\sqrt{I}$ nm with I in mol/L, and the usable envelope is roughly $3\kappa^{-1}$.

A working Bst 2.0 buffer runs at roughly 50 mM monovalent salt plus 10 mM Mg^{2+}, so $I \approx 0.09$ M. At the B-form contour rise of 0.34 nm/bp and vertical anchoring, the envelope converts directly into a base count (Table 7).

The operating range is therefore roughly 9 to 26 bases, with buffer ionic strength as the tunable and polymerase activity as the floor. Two obvious levers both fail at fixed sensing chemistry. Extending the depth to reach 500 bp means 170 nm of active-site travel, requiring $\kappa^{-1} \approx 55$–85 nm and therefore $I \approx 26\ \mu$M, effectively deionized water, in which Bst 2.0 does not turn over at all, since two-metal-ion phosphoryl transfer requires millimolar Mg^{2+}. Flattening the strand to fit 500 incorporations inside a 9 nm envelope requires an average rise below 0.018 nm/bp, about 5% of contour, holding the duplex within about three degrees of the surface plane for all 500 bases; the ~50 nm persistence length of dsDNA guarantees excursions exceeding 9 nm long before base 500.

One candidate lever does *not* help. Operating at 65 °C does not extend the envelope: $\kappa^{-1} \propto \sqrt{\varepsilon_r T}$, and the rise in T from 298 to 338 K is very nearly cancelled by the fall in water's relative permittivity from about 78 to 64, leaving $\kappa^{-1}(65\,^\circ\mathrm{C}) \approx 0.98\,\kappa^{-1}(25\,^\circ\mathrm{C})$. Temperature is not a lever.

F.2 Extending the Sensing Horizon

The Debye length is not a constant of the system, and treating the horizon above as a wall would misstate the literature. κ^{-1} derives from the linearized Poisson–Boltzmann model, whose assumptions of small surface potentials, thermodynamic equilibrium, and point-charge ions with no steric crowding are all violable by design, and a substantial body of work reports label-free electrical detection at physiological ionic strength by doing so [55].

Four approaches are established enough to name. Dense polyelectrolyte multilayers raise the entropic cost of confining ions inside the film and have been reported to extend the effective screening length from 0.8 to about 10 nm at physiological ionic strength [56]. Biomolecule-permeable polymer coatings achieve detection in 150 mM buffer where unmodified devices fail above 10 mM [57,58]. Non-equilibrium high-frequency operation drives the interface faster than the ion cloud can respond; a CMOS nanocapacitor array architecturally close to ECSEQ-0's own has shown sensing depth extending past κ^{-1} as measurement frequency rises toward 50 MHz [34,35], a readout-electronics lever rather than a chemistry one, developed above. Finally, concave electrode topography restricts the volume in which double layers can form [59]; ECSEQ-0's planar array is the worst case on that axis, and recessed pixel geometry is an unexplored, fabrication-side option.

What has *not* been demonstrated, anywhere we can find, is any of this over a catalytically active enzyme. Every result above was obtained on a static binding event, in monovalent buffer or at most 2 mM Mg^{2+}, at room temperature. ECSEQ-0 needs 10 mM Mg^{2+} at 65 °C with a strand-displacing polymerase and a growing duplex moving through whatever layer is doing the screening extension, and divalent cations screen far more effectively per mole and can collapse polyelectrolyte films, while polymer brushes can dehydrate on heating and lose the volume fraction the mechanism depends on. Nor is there data on such a layer surviving 50–100 regeneration cycles. This is the largest unquantified risk in the platform.

The nearest precedent shows the size of the gap. Single-molecule electronic recordings of DNA polymerase I processing have been obtained in 10 mM Tris / 50 mM NaCl / 10 mM $MgCl_2$, essentially ECSEQ-0's

Table 8: Per-cycle timing of the incorporation sub-steps. The chemistry variance is the within-base value at the mean rate, which is the quantity that dephases copies of one template; between-base spread is signal, not jitter.

Sub-step	Distribution	Mean	Variance
Search	Exp(0.10 ms^{-1})	10 ms	100 ms^2
Chemistry	Gamma(k = 2.78), base-dependent rate	66–116 ms (mean 87.5)	2754 ms^2
Translocation	Exp(0.05 ms^{-1})	20 ms	400 ms^2
Cycle		117.5 ms	3254 ms^2

ionic strength, resolving individual incorporations over more than 10,000 bond-forming events [60]. That work does not discuss screening at all, because the enzyme is conjugated directly to the transducer and the standoff distance is approximately zero. The engineering problem this paper poses is therefore not to defeat the Debye limit; it is to reproduce that demonstrated result at a standoff of a few nanometres rather than none.

Bounded target. The target is far smaller than the read-length ceiling suggests. A read does not need to be long; it needs to be *mappable*. Unique placement against a 3.2 Gb reference needs roughly 20–25 bases, so the objective is to extend the effective sensing horizon from its nominal ~9 bases to that floor, a factor of ~3, not the three orders of magnitude that would put ECSEQ-0 in competition with translocation sequencing on read length. A 3× extension is a tractable engineering bet; matching nanopore read lengths is not, and pursuing it would trade away the architecture's differentiation. The kill-criterion is measurable: if single-pixel validation cannot extend the horizon across the mapping floor at a polymerase-compatible ionic strength, the whole-genome application is closed and the platform remains a targeted-panel device.

F.3 Phase-Spread Model

Dephased copies do not vanish. Leading and lagging strands continue to contribute signal at the wrong position, appearing as inter-symbol interference rather than a clean loss of amplitude. The faithful account models each strand's incorporation index as a random walk: copies drift independently, so the ensemble's positional distribution has standard deviation

$$\sigma_n = \sqrt{D\,n}\ \text{bases} \tag{6}$$

with D the per-cycle phase-step variance (in $base^2$). Contrast between adjacent positions survives only while σ_n stays below roughly one base, giving $N_{\max} \approx 1/D$.

D follows from the per-copy cycle-time distribution, which the simulator's kinetics configuration already fixes. One cycle is three sequential sub-steps (Table 8).

The cycle mean reproduces the $t_{\text{base}} = 0.1175$ s used throughout the throughput analysis exactly, so the timing model is self-consistent; it was only the dephasing model that was not. The per-cycle standard deviation is $\sigma = \sqrt{3254} = 57.0$ ms, a coefficient of variation $CV = 57.0/117.5 = 0.485$. A copy advances one base per cycle, so the CV is the positional step jitter measured in bases:

$$\begin{aligned} D &= \mathrm{CV}^2 = 0.235\ \text{base}^2/\text{cycle} \\ \sigma_n &= 0.485\sqrt{n}\ \text{bases} \end{aligned} \tag{7}$$

σ_n crosses one full base at $n \approx 4$ (Fig. 13); by $n = 100$ it is 4.9 bases, and by the ~1,150 cycles a single-run 30× genome would need at panel scale it is 16 bases. The criterion $N_{\max} \approx 1/D$ therefore gives $N_{\max} \approx 4$ on the nominal kinetics, two orders of magnitude below the attrition bound.

The chemistry sub-step carries 85% of the cycle variance and is the obvious target, but tightening it does not recover a long read. Make chemistry perfectly deterministic, removing all 2754 ms^2, and the two exponentially distributed sub-steps alone still leave 500 ms^2, $\sigma = 22.4$ ms, $CV = 0.190$, and a one-base crossing at $n \approx 28$. Read length is bounded to tens of bases by search and translocation regardless of how tightly the chemistry step is controlled. Extending it is a question of changing the kinetics themselves, an engineered polymerase whose incorporation cycle is composed of many low-variance steps, not of tightening a tolerance on the current one.

F.4 Convergent Ceilings

Phase spread is not the only mechanism pointing at a short read, and the others are physically independent of it. The Debye screening length at the operating ionic strength bounds the distance over which a growing strand stays visible to the electrode at all, and the $1/n$ dilution of one incorporation's contribution against an n-base duplex already synthesized erodes the marginal signal as the read extends. Both land in the same range of roughly ten to a few tens of bases.

The measurement that decides it is the per-copy cycle-time distribution for Bst 2.0 at 65 °C in ECSEQ-0's buffer, specifically its coefficient of variation rather than its mean, since everything above is a function of CV alone. It is a single-pixel, or even a bulk single-molecule

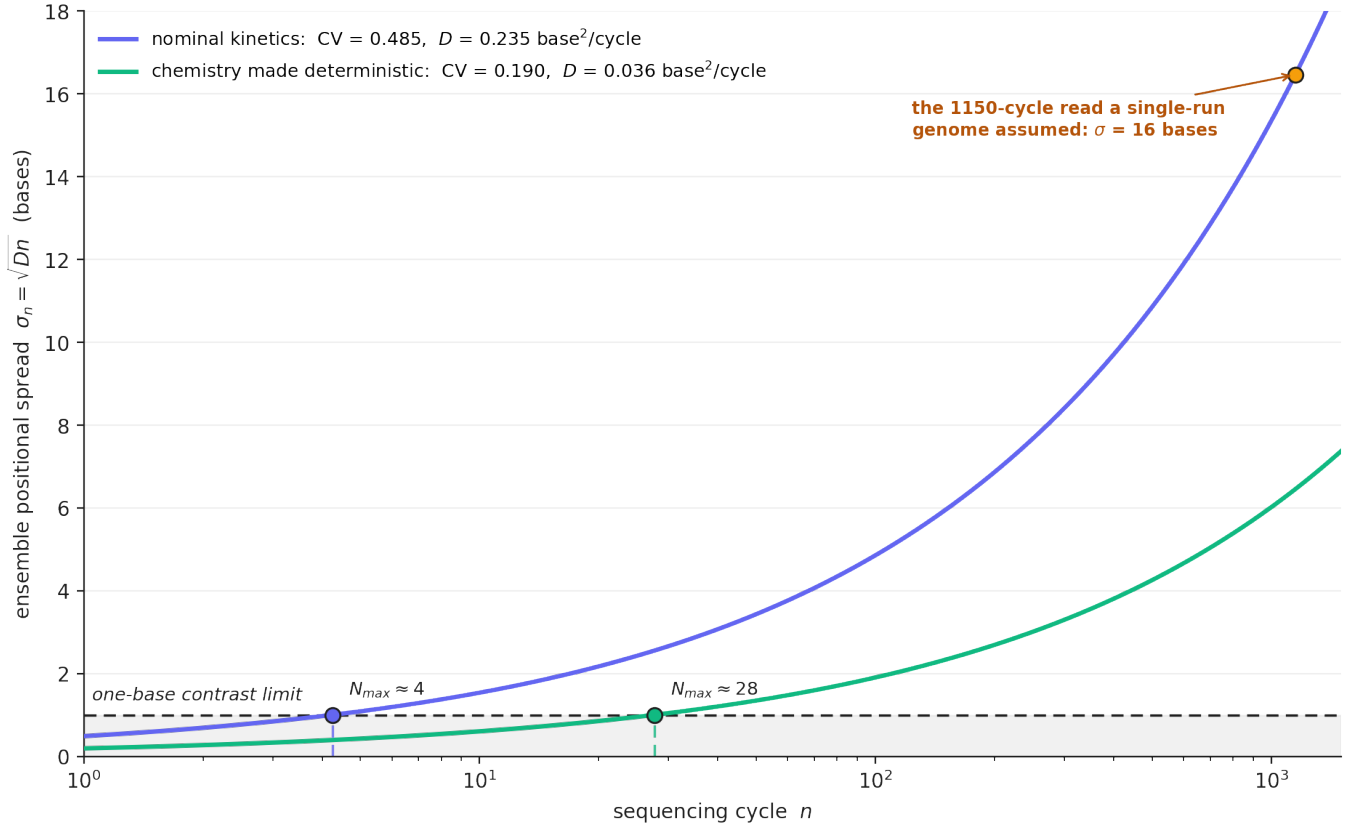

Fig. 13: Positional spread of the ensemble, $\sigma_n = 0.485\sqrt{n}$ bases, against cycle number; adjacent positions stop resolving once σ_n exceeds one base.

kinetics, measurement, and it sits at the top of the valida- tion list.