Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

How much of a finite sample can a microchannel surface sensor capture? Optimal flow rate, universal capture map and closed-form design rules

Ananya Sharma

ananya.sharma.pune@gmail.com

MIT Art, Design and Technology University

Research Article

Keywords: Mass transport, Surface-based biosensor, Heterogeneous immunoassay, Sample volume, Damköhler number, Point-of-care diagnostics

Posted Date: November 1st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11390083/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Surface-capture immunoassays and biosensors in microchannels are usually analysed as if the sample never runs out. A diagnostic sample does: a few microlitres must be processed within a fixed assay time, and the detection limit is set by how many analyte molecules reach the capture surface. I modelled a capture patch on the floor of a rectangular channel through which a finite sample is pumped, using a steady marching solver and an independent two-dimensional transient solver with Langmuir kinetics. Both were validated against the Lévêque, Graetz and reaction-limited solutions and against each other to within 0.5%. The flow rate that maximises capture is the one that delivers the whole sample in exactly the available time, Q* = V/T, and transient simulations confirmed this in every case tested, including near saturation. At that flow rate the captured fraction depends on only two dimensionless groups and is bounded by a transport ceiling and a kinetic ceiling. The kinetic group contains neither channel height nor diffusivity. A closed-form combination of the two ceilings reproduces the numerical solution to within − 2.2% to + 1.4% over the whole design space. Recirculation helps only up to the kinetic ceiling. For a nominal point-of-care design, pumping ten times faster than Q* cut the captured fraction from 8.5% to 1.2% and raised the detection limit from 3.9 fM to 27.6 fM. Simple design rules follow, and the code is provided.

1 Introduction

A surface-capture assay in a microchannel is a race between two clocks. Analyte molecules have to diffuse from the bulk of the passing sample to an antibody-coated patch on the floor, and they have to do it before they are swept out of the channel. Squires, Messinger and Manalis laid out the physics of that race for surface biosensors: the Lévêque solution for the collection flux, the full-collection regime at low Péclet number, and the competition between transport and reaction (Squires et al. 2008). A long line of numerical work had already solved the convection–diffusion–reaction problem in detail, mostly with kinetic-constant extraction in mind (Glaser 1993; Myszka et al. 1998; Vijayendran et al. 1999; Gervais and Jensen 2006; Hu et al. 2007).

Almost all of that analysis assumes the sample never runs out. In a kinetics experiment on a commercial optical biosensor that is a fair assumption: the pump keeps delivering fresh analyte at a fixed concentration (Sjoelander and Urbaniczky 1991). A diagnostic assay is different. A finger-prick or a few microlitres of plasma is all there is, and the clinician wants an answer within a set time. Under those two constraints the useful question is no longer "how fast do molecules bind?" but "what fraction of the molecules in my sample end up on the sensor by the time I read it?" At low concentrations that fraction sets the detection limit directly, because a readout needs a minimum number of bound molecules regardless of how they got there (Sheehan and Whitman 2005; Nair and Alam 2006). The pressure is sharpest for blood biomarkers of neurodegeneration. Neurofilament light chain and phosphorylated tau 217 circulate in plasma at picomolar to sub-picomolar levels (Khalil et al. 2018; Palmqvist et al. 2020), which is why single-molecule methods that count individual captured proteins were needed to measure them at all (Rissin et al. 2010).

The finite-volume problem is not new, and I want to be clear about what was known before. Parsa et al. (2008) studied it computationally and experimentally. They identified a "reagent-limited" regime that exists only when sample volume and assay time are both fixed, reported a critical flow rate that gave the maximum signal in the shortest time for their device, and showed that a short concentrated plug beat a long dilute one carrying the same number of molecules. Zimmermann et al. (2005) optimised flow velocity in a finite-difference model of a low-volume surface immunoassay. Chou et al. (2012) found an optimal flow rate for porous-bead sensors under volume and time limits and showed that recirculating the sample recovers signal. Static microspot assays approach the same accounting from the other side: Ekins (1989, 1998) and Saviranta et al. (2004) showed that a small spot removes only a tiny fraction of the analyte in its well. Others have since pushed toward capturing nearly everything, with large-volume processing in paper (Miller et al. 2018) or nanofluidic total capture (Galvin et al. 2018).

What I could not find, after the searches summarised in Online Resource 1, is a general answer. Each optimum above was found for one geometry and one chemistry, by running a simulation for that case. No study I found gives the condition under which the optimum flow rate is simply the one that delivers the sample over exactly the available time. None reduces the best achievable captured fraction to a map in dimensionless groups, or gives a closed form for it that a designer could evaluate on paper. And none bounds how much recirculation can help.

This paper tries to provide those pieces. I treat a rectangular microchannel with a capture patch spanning its floor, a fixed sample volume V and a fixed assay time T. Section 2 sets up the model, two independent solvers and their validation. Section 3 shows that the optimum is Q* = V/T whenever the per-pass capture fraction is concave in the Graetz variable, which it always was in my calculations. It then maps the captured fraction at that optimum over two groups, gives a closed form and a kinetic ceiling, quantifies the penalty for running too fast, and converts the result into detection limits. Section 4 turns this into rules a person designing a chip can actually use, and lists what the model leaves out. Everything is computational. No experimental data were generated, and the biomarker numbers appear only to give the arithmetic a realistic scale.

2 Methods

2.1 Geometry and governing equations

The channel has height H and width W, with W much larger than H, so the flow is treated as two-dimensional between parallel plates. A capture patch of length L covers the full width of the floor. A sample of volume V and analyte concentration c_0 is pumped through at volumetric flow rate Q and is followed by analyte-free buffer, and the sensor is read at time T after the sample first enters. The velocity profile is fully developed plane Poiseuille flow, $u(y) = 6U(y/H)(1 - y/H)$, with mean velocity $U = Q/(WH)$. Analyte transport obeys

$$\partial c/\partial t + u(y)\, \partial c/\partial x = D\, (\partial^2 c/\partial x^2 + \partial^2 c/\partial y^2), \quad (1)$$

with no flux through the walls except at the patch, where binding follows Langmuir kinetics,

$\partial b/\partial t = k_{on}\, c_w\, (b_m - b) - k_{off}\, b = D\, \partial c/\partial y$ at $y = 0$. (2)

Here b is the surface density of bound analyte, b_m the density of capture sites, c_w the concentration at the wall, and k_{on} and k_{off} the association and dissociation rate constants. The diffusion coefficient D was estimated from molecular mass with the correlation of Young et al. (1980) at 25°C in water.

2.2 Dimensionless groups and the finite-sample accounting

In the low-occupancy limit ($b \ll b_m$) Eq. (2) is linear, and over the sensor the steady problem depends on only two groups. One is the Graetz variable $X = DWL/(QH)$, the ratio of the residence time over the patch to the time to diffuse across the channel. The other is the Damköhler number $Da = k_{on}\, b_m\, H/D$, the ratio of the reaction rate at the wall to the diffusive supply. I define the per-pass capture fraction $G(X, Da)$ as the fraction of the analyte flux entering the channel that binds to the patch at steady state.

If the transport and binding transients are short compared with T, and the channel holds a small fraction of V, the number of captured molecules is

$N = c_0\, V\, G(X)$ for $Q \geq V/T$ (the whole sample passes within T),

$N = c_0\, Q\, T\, G(X)$ for $Q < V/T$ (only the volume QT passes). (3)

The captured fraction $f = N/(c_0\, V)$ therefore depends on Q only through X and through the factor $\min(QT/V, 1)$. Evaluating X at the flow rate $Q^* = V/T$ gives the group

$\Omega = DWLT/(HV)$, (4)

and the product $\kappa = \Omega \cdot Da = k_{on}\, b_m\, W\, L\, T/V$, which contains neither D nor H. Section 3.2 shows that κ sets a kinetic ceiling on capture.

I checked the quasi-steady assumption behind Eq. (3) directly with the transient solver (Section 2.4), which imposes the sample as a finite plug followed by buffer and makes no such assumption.

2.3 Steady marching solver for G(X, Da)

Neglecting axial diffusion (valid when the Péclet number $UH/D \gg 1$; at the nominal optimum it is about 540), the steady form of Eq. (1) is parabolic in x and can be marched downstream. In scaled variables $Y = y/H$ and X it reads $(u/U)\, \partial c/\partial X = \partial^2 c/\partial Y^2$, with $\partial c/\partial Y = Da \cdot c$ at $Y = 0$ and $\partial c/\partial Y = 0$ at $Y = 1$. I discretised Y with a cell-centred finite-volume grid, geometrically stretched toward the patch (first cell 2×10^{-5} H, 120 cells), and marched in X on a geometric grid of 3000 steps. The first two steps used backward Euler and the rest variable-step second-order backward differentiation. The wall flux uses a half-cell resistance, so the Robin condition is exact for a linear profile in the first cell. G is accumulated as the

integrated wall flux and checked against the depletion of the flux-weighted concentration; the two agree to round-off. A table of G was computed on 131 values of X between 10^{-5} and $10^{1.5}$ and 25 values of Da between 10^{-3} and 10^{3}, plus the fully absorbing limit Da → ∞.

2.4 Transient two-dimensional solver

To test everything the marching solver assumes, I wrote a second solver in dimensional form. It integrates the full Eq. (1), with axial and transverse diffusion, on a stretched finite-volume grid covering half a patch length upstream and a quarter downstream of the patch. Advection is first-order upwind, time stepping is backward Euler, and the nonlinear Langmuir condition (2) is applied cell by cell at the patch, with the bound density updated implicitly. The inlet concentration can follow any time course. I used a plug of duration V/Q followed by buffer, or a well-mixed reservoir that receives the outlet stream and feeds the inlet, to model recirculation. Mass balance (analyte in = analyte out + in the channel + bound) was tracked at every step.

2.5 Closed-form estimate

Two limits of f* = G(Ω, Da) are exact. When binding is infinitely fast (Da → ∞), f* equals the transport ceiling $G_{\infty}(\Omega)$, the capture fraction of a perfectly absorbing patch. When transport is infinitely fast, the patch sees a well-mixed sample and f* = 1 − exp(−κ). I combined the two as transfer units in series, by analogy with heat exchanger design:

$$f^{*} \approx 1 - \exp(-N_{tot}),\ 1/N_{tot} = 1/N_{t} + 1/\kappa,\ N_{t} = -\ln[1 - G_{\infty}(\Omega)]. \quad (5)$$

G_{∞} itself has a fully explicit approximation: the smaller of the Lévêque asymptote $G \approx 1.467\ \Omega^{2/3}$, unity, and the fully developed Graetz decay $1 - C \exp(-\mathrm{Nu}\cdot\Omega/2)$. Nu = 4.8608 is the Nusselt number for parallel plates with one wall at uniform concentration and the other insulated (Shah and London 1978), and C is a constant fitted once to the marching solution. A simpler harmonic combination, $1/f^{*} = 1/G_{\infty} + 1/(1 - e^{-\kappa})$, is reported for comparison.

2.6 Nominal design and parameter ranges

The nominal device (design A in Table 1) is a channel 50 μm high and 1 mm wide with a 1 mm × 1 mm capture patch. It processes 50 μL of sample in 15 min. The analyte is a 20 kDa protein ($D = 1.03 \times 10^{-10}$ $m^2\ s^{-1}$), captured by an antibody with $k_{on} = 10^{6}\ M^{-1}\ s^{-1}$ and $k_{off} = 10^{-4}\ s^{-1}$, typical of affinity-matured antibodies (Foote and Eisen 1995), at a site density of $b_m = 10^{-8}$ mol m^{-2} (about 6000 sites per μm^2). Eight variants change one or two parameters each (Table 1). Single-parameter sweeps covered H from 10 to 200 μm, L from 0.1 to 10 mm, molecular mass from 5 to 500 kDa, k_{on} from 10^{4} to $10^{7}\ M^{-1}\ s^{-1}$, b_m from 10^{-10} to 10^{-7} mol m^{-2}, V from 5 to 100 μL and T from 5 to 120 min.

To translate capture into a detection limit I assume the readout needs N_{min} bound molecules, so that $c_{LOD} = N_{min}/(N_A\ f\ V)$, where N_A is the Avogadro constant. N_{min} was varied from 10^{3} to 10^{5}. This is a

deliberately readout-agnostic choice: the right value depends on the label, the imaging and the background, none of which the model contains.

2.7 Software, reproducibility and use of AI tools

All calculations used Python 3.13 with NumPy (Harris et al. 2020), SciPy (Virtanen et al. 2020), pandas and Matplotlib (Hunter 2007). The calculations are deterministic. One script regenerates every table, figure and number in this paper in about nine minutes on a laptop, and records the software versions it used. Figures use the Okabe–Ito colour-blind-safe palette, with line styles and markers as a second channel.

I used a large language model, Claude Opus 5.5 (Anthropic), as an assistant during this work. It helped debug the simulation and plotting code, run and screen the literature and novelty searches and check reference metadata against Crossref. I specified the problem and the model, checked the code against the analytic limits reported in Section 3.1, verified every reference against its publisher record, and checked every number in the text against the output files. I take full responsibility for writing the content.

3 Results

3.1 The two solvers agree with each other and with the analytic limits

Before using either solver I checked them against every limit I could solve by hand (Fig. 1). For a perfectly absorbing patch at small X the marching solver reproduces the Lévêque asymptote $G = 1.467\ X^{2/3}$. The ratio to the asymptote lies between 0.986 and 0.999 for $X \le 10^{-3}$ and drifts below one only as the depletion layer starts to feel the opposite wall (Fig. 1a). Far downstream, $1 - G$ decays exponentially at a rate that corresponds to a Nusselt number of 4.861, against the tabulated 4.8608 (Shah and London 1978), a difference of 0.014% (Fig. 1b). In the reaction-limited regime ($Da = 10^{-3}$ and 10^{-2}) G matches the well-mixed result $1 - \exp(-Da \cdot X)$ to within 0.36%. Refining the grid at the nominal point and at two transport-limited points gave an observed convergence order of 2.01, as expected for the scheme, and the finest grid changed G by less than 6×10^{-5} in relative terms.

The transient solver reproduced the Langmuir binding curve of a reaction-limited, saturating case to within 0.7% over 400 s (Fig. 1c). At steady state it agreed with the marching solver to within 0.35–0.48% on its finest grid of 83,904 cells, with the gap closing as the grid was refined (Fig. 1d). The residual reflects the axial diffusion and first-order upwinding that the transient solver includes and the marching solver omits. Mass was conserved to better than 5×10^{-14} of the analyte delivered in every run.

3.2 The best flow rate is the one that just delivers the sample in time

Figure 2a shows the depletion layer that grows over the patch at the nominal optimum: most of the channel passes over the sensor untouched. Figure 2b plots the captured fraction against Q/Q* for three channel heights. Symbols are full transient simulations, with the sample delivered as a finite plug followed by buffer and read at time T. Lines are the quasi-steady prediction of Eq. (3). In all three cases the transient optimum fell exactly at Q = Q* = V/T. The same held for a 10 µL sample, and for a 1 nM sample in which the capture sites approach saturation and the nonlinear Langmuir term is active. In the linear cases the quasi-steady formula overestimated the transient result by 2–8% across the whole range of flow rates. In the saturating 1 nM case it overestimated it by 8–15%, which is expected, since Eq. (3) uses the low-occupancy capture fraction. The overestimate arises because the transient solver accounts for the time the sample needs to reach the patch and for the analyte left in the channel at readout.

The reason the optimum sits at V/T is simple once Eq. (3) is written down. Above Q* the whole sample passes, so the only effect of pumping faster is a shorter residence time, and capture falls. Below Q* the sample is not all used. The number captured is then the capture rate $c_0 \cdot Q \cdot G$ times the time T, and this rises with Q as long as Q·G does. In terms of X, that requires $d \ln G / d \ln X \le 1$. The condition holds whenever G is concave in X with G(0) = 0, and I checked it numerically on the full table: the largest local slope of ln G against ln X was 0.99999, reached in the reaction-limited corner where G is proportional to X. So within this model there is no case where it pays to leave sample unused, and no case where it pays to rush it. In the reaction-limited corner the curve below Q* is nearly flat, and the optimum is weak rather than sharp.

3.3 A universal capture map, a kinetic ceiling and a closed form

Because the optimum is always at Q*, the best achievable captured fraction is f* = G(Ω, Da), a function of two groups only. Figure 3a maps it. The map has two walls. Moving right (larger Ω: longer patch, longer assay, shallower channel, smaller sample) lowers the transport wall. Moving up (larger Da: more sites, faster binding) lowers the kinetic wall. The dashed lines are contours of $\kappa = \Omega \cdot Da$. Along a line of constant κ, f* cannot exceed $1 - \exp(-\kappa)$, the fraction a perfectly mixed sample would lose to the patch in time T. The bound was never violated anywhere in the table. Note what $\kappa = k_{on}\, b_m\, W\, L\, T/V$ does not contain: the channel height and the diffusion coefficient. Making the channel shallower can only bring a design up to its kinetic ceiling, never past it.

Equation (5) reproduced f* over the 3086 tabulated points with relative errors between − 2.2% and + 1.4% and a median absolute error of 0.15% (Fig. 3c). With the fully explicit approximation for G_∞, so that no table is needed at all, the error range widened to − 2.1% to + 6.4% with the same median. The worst errors sit at Ω near 1, where neither the Lévêque nor the Graetz approximation of G_∞ is accurate. The fitted Graetz constant was C = 0.898. The harmonic combination, by contrast, underestimated f* by up to 50% once capture exceeded half the sample, because adding resistances in that form ignores how depletion compounds along the patch.

Table 1 places nine example designs on the map. The nominal device sits at Ω = 0.037 and Da = 4.9 and captures 8.5% of the sample at Q*. Its kinetic ceiling is 16.5%.

Table 1
Example designs at Q* = V/T. Nominal design A: H = 50 µm, W = 1 mm, L = 1 mm, V = 50 µL, T = 15 min, 20 kDa analyte, k_{on} = 10^6 M^{-1} s^{-1}, k_{off} = 10^{-4} s^{-1}, b_m = 10^{-8} mol m^{-2}. Designs B–I differ from A only in the parameters listed. LOD: lowest detectable concentration if the readout needs 10^4 bound molecules

Design	Change from A	Ω	Da	κ	Q* (µL min^{-1})	f* (%)	f at 10Q* (%)	f, 20 passes (%)	LOD at Q* (fM)	LOD at 10Q* (fM)
A	nominal	0.037	4.9	0.18	3.33	8.5	1.2	12.1	3.9	27.6
B	H = 20 µm	0.093	1.9	0.18	3.33	11.2	1.4	13.8	3.0	23.4
C	H = 100 µm	0.019	9.7	0.18	3.33	6.5	1.0	10.5	5.1	33.0
D	L = 5 mm	0.19	4.9	0.90	3.33	29.7	4.8	41.4	1.1	6.9
E	V = 10 µL	0.19	4.9	0.90	0.667	29.7	4.8	41.4	5.6	34.5
F	T = 60 min	0.15	4.9	0.72	0.833	25.2	4.0	35.7	1.3	8.3
G	150 kDa analyte	0.019	9.5	0.18	3.33	6.6	1.0	10.5	5.1	32.8
H	k_{on} = 10^5 M^{-1} s^{-1}	0.037	0.49	0.018	3.33	1.6	0.17	1.7	20.4	193
I	H 20 µm, L 5 mm, T 30 min	0.93	1.9	1.8	1.67	64.9	11.2	70.5	0.51	3.0

3.4 Running too fast is expensive, running slightly slow is cheap

Figure 3b shows what happens away from Q*. Above it, the captured number falls roughly as $(Q/Q^*)^{-2/3}$, the Lévêque scaling, and somewhat faster when binding also limits capture. At the nominal design, delivering the sample ten times faster than Q* (in 1.5 min instead of 15) leaves only 14% of the optimal capture: a seven-fold loss in signal from a choice that costs nothing to get right. The useful tolerance is narrow on the fast side. For Ω ≤ 0.1, which covers most of the designs in Table 1, keeping at least 90% of the optimum requires Q to stay below 1.09–1.15 Q*. The band widens only for designs that already capture most of the sample (up to 1.45 Q* at Ω = 1 and to 10.9 Q* at Ω = 10 with fast binding). Below Q* the loss is gentler for small Ω, because the capture rate per unit flow rises when the flow slows (the Ω =

0.037 curve in Fig. 3b). The practical reading is to aim for the sample to finish entering the channel just before readout, and to err slow rather than fast.

3.5 Which design levers move the optimum

Figure 4 varies one parameter at a time from the nominal design and separates the full model from its two ceilings. Channel height is a weak lever here. Going from 200 μm to 10 μm raised f* from 4.7% to 12.9%, under a factor of three over a twenty-fold change in height, because the design hits its kinetic ceiling (Fig. 4a). Analyte size matters even less: from 5 to 500 kDa, f* fell only from 9.9% to 5.5%. The strong levers are the ones inside κ. Patch length took f* from 1.2% at 0.1 mm to 48% at 10 mm (Fig. 4b). The binding rate constant took it from 0.18% at 10^4 M^{-1} s^{-1} to 14% at 10^7 M^{-1} s^{-1}, with the curve following the kinetic ceiling almost exactly until the transport ceiling takes over (Fig. 4d). Capture-site density acted identically to k_{on}, as it must, since the two appear only as their product. Sample volume and assay time enter through Ω and κ together: f* rose from 4.8% to 48% as the volume fell from 100 to 5 μL, and from 3.4% to 42% as the assay time rose from 5 to 120 min.

Combining the two strongest levers that do not depend on chemistry (a 20 μm channel and a 5 mm patch, with 30 min of assay time, design I) raised capture of the 50 μL sample to 65%. For comparison, capturing half the sample with the nominal geometry would take 158 min at its own Q*.

3.6 Recirculation helps only up to the kinetic ceiling

If the outlet is fed back to a well-mixed reservoir, each molecule gets several chances to bind. With q = QT/V passes the captured fraction becomes $1 - \exp[-q \cdot G(\Omega/q, Da)]$. The transient solver, run with an explicit recirculating reservoir, gave 10.2% and 11.7% at 5 and 20 passes. The formula gave 10.4% and 12.1%, about 3% higher, because the formula ignores the volume held up in the channel (Fig. 5a). For the nominal design recirculation lifts capture from 8.5% to 12.1% at 20 passes and to 15.0% at 1000. It can never pass the kinetic ceiling of 16.5%, since a perfectly recirculated sample is a perfectly mixed one. When binding is slow the ceiling sits right on top of the single-pass value: with k_{on} = 10^5 M^{-1} s^{-1} recirculation moves capture from 1.6% to at most 1.8%. Recirculation is a large lever only when transport is the bottleneck. In the transport-limited case it took capture from 15% to 80%.

3.7 What this means for a detection limit

If the readout needs N_{min} bound molecules, the lowest detectable concentration is $N_{min}/(N_A\, f\, V)$, so every factor in f carries straight through to the detection limit (Fig. 5b). With N_{min} = 10^4, the nominal device detects 3.9 fM at Q*, but 27.6 fM if the sample is pushed through ten times too fast (Table 1). The combined design I reaches 0.51 fM, and the slow binder (design H) only 20 fM. To put this on a clinical scale, plasma neurofilament light and p-tau217 sit at low-picomolar and sub-picomolar concentrations (Khalil et al. 2018; Palmqvist et al. 2020). The difference between running at Q* and running at 10Q* is therefore the difference between a comfortable margin and a marginal measurement, without any change in chemistry or readout.

3.8 Consistency with published experiments

Two published experimental trends fall out of the model without any fitting. Parsa et al. (2008) found that a short concentrated plug gave more signal than a long dilute plug carrying the same number of molecules. In the model, a 5 µL plug delivered at its own Q* captured 2.8 to 9.5 times more molecules than a 50 µL plug with ten-fold lower concentration, across channel heights of 20–100 µm, assay times of 5–60 min and two binding rate constants. The ratio approaches ten only when capture is far from both ceilings. Lynn et al. (2013) showed that the analyte flux scales as $H^{-2/3}$ at fixed flow rate and measured a four-fold improvement in detection limit when the chamber height was reduced from 47 to 7 µm. For an absorbing patch at a fixed 10 µL min^{-1} the marching solver gives a 3.42-fold increase in capture over the same change, close to the pure Lévêque ratio of 3.56. This is a consistency check on the transport physics, not a quantitative validation against either study. I did not have their full device parameters, and their experiments were not run under a fixed-volume constraint.

4 Discussion

4.1 What is new here and what is not

The existence of an optimum flow rate for a finite sample was shown before, by simulation and by experiment, for specific devices (Zimmermann et al. 2005; Parsa et al. 2008; Chou et al. 2012). What this paper adds is the general version. The optimum is Q* = V/T whenever per-pass capture is concave in the Graetz variable, which held over the entire range I computed. The best achievable fraction depends on two groups, Ω and Da. And it is bounded and well approximated by two ceilings that can be written down on paper: a transport ceiling set by Ω alone and a kinetic ceiling set by $\kappa = k_{on}\, b_m\, W\, L\, T/V$ alone. The closed form in Eq. (5) is accurate to a few percent, which is better than the uncertainty in k_{on} or b_m in almost any real assay. So a designer can place a proposed chip on the map with a calculator before building or simulating anything.

4.2 Practical design rules

These follow directly from Sections 3.2–3.7.

1. 1. Set the flow rate so the sample finishes entering the channel just before readout: Q* = V/T. Err slow, not fast. For Ω ≤ 0.1, exceeding Q* by 10–15% already costs 10% of the signal, and a ten-fold overshoot costs 77–90%.
2. 2. Compute $\kappa = k_{on}\, b_m\, W\, L\, T/V$ first. If 1 − exp(−κ) is below the capture fraction you need, no channel geometry, mixer or recirculation scheme will get you there. Only a longer or wider patch, more or faster capture sites, more time or less volume will.
3. 3. Compute Ω = DWLT/(HV) next. If $G_\infty(\Omega)$ (about $1.47\,\Omega^{2/3}$ for $\Omega \lesssim 0.1$) is far below the kinetic ceiling, transport is the bottleneck. Then a shallower channel, a mixer (Stroock et al. 2002; Foley et

al. 2008; Lynn et al. 2014), hydrodynamic focusing of the sample onto the sensor (Hofmann et al. 2002) or recirculation are worth adding. If the two ceilings are close, none of these will help much.

4. 4. Use Eq. (5) for the estimate and the map in Fig. 3a to see which wall you are against. Then convert to a detection limit with $c_{LOD} = N_{min}/(N_A f^* V)$.
5. 5. When a fixed number of molecules can be delivered as a smaller, more concentrated sample, do it. Smaller V raises Ω and κ together.

Rule 2 runs against a common intuition. In the nominal design, cutting the channel height from 50 to 20 µm, which is often the first change people make, improved capture by only about a third, from 8.5% to 11.2%. Making the patch five times longer more than tripled it, to 29.7%. Nothing about this is surprising once κ is on the page, but it is easy to miss when the analysis starts from the Lévêque flux, which is all about height (Squires et al. 2008; Lynn et al. 2013).

4.3 Relation to other assay formats

The same accounting applies, with different transport functions, to flow-through formats in which the sample passes through rather than over the capture surface: porous beads (Chou et al. 2012), paper (Miller et al. 2018), nanohole arrays (Escobedo et al. 2010) and nanochannels that capture nearly everything (Galvin et al. 2018). There the effective Ω is much larger because the diffusion distance is much shorter. Finite-sample depletion has also been modelled for static chambers (Jokić et al. 2021), and in microarray spots mass transport sets both kinetics and the regime of operation (Kusnezow et al. 2006; Zhao et al. 2010). In the ambient-analyte limit, where a tiny spot removes almost nothing from its sample (Ekins 1989; Saviranta et al. 2004), f is small by design and the signal reports concentration, not amount. Flow-through capture of a finite sample sits at the opposite end: the signal reports how many molecules were caught, which is why the captured fraction is the quantity to optimise. Calibration-based analyses of competitive assays face related trade-offs (Fu et al. 2009), and analyte concentration can be extracted from transport-limited binding on optical biosensors (Christensen 1997).

4.4 Limitations

The model is two-dimensional and treats the patch as spanning the full width. Side-wall regions of a real channel, where the velocity drops, and patches narrower than the channel will change G somewhat, though not the structure of the argument. The quasi-steady accounting in Eq. (3) needs the channel volume to be small compared with the sample and the transport transient to be short compared with T. The transient simulations confirm both for the designs here, with differences of 2–8% in the linear regime, but neither assumption holds for a device whose dead volume is comparable to V. Near saturation the linear formula overestimates capture further (by up to 15% in the 1 nM case), although the optimum stayed at Q*. The detection-limit arithmetic ignores nonspecific binding, matrix effects in plasma, surface heterogeneity of the capture layer (Vijayendran et al. 1999) and the noise model of any particular readout. N_{min} stands in for all of them. The surface reaction is first-order Langmuir with uniform sites. Finally, and most importantly, this is a computational study. I compared it with published

experimental trends but did not run experiments, and the design rules should be read as predictions to test, not as measured performance.

5 Conclusions

For a surface-capture assay with a fixed sample volume and a fixed assay time, the flow rate that captures the most analyte is the one that delivers the sample over exactly the available time, $Q^* = V/T$. The captured fraction at that flow rate depends on two dimensionless groups and is bounded by a transport ceiling and a kinetic ceiling. A closed-form combination of the two ceilings reproduces the full numerical solution to within a few percent. The kinetic group $\kappa = k_{on}\, b_m\, W\, L\, T/V$, which does not involve channel height, sets a hard limit that neither geometry nor recirculation can beat. At typical point-of-care parameters, running a sample ten times faster than Q^* threw away 86% of the achievable signal in the nominal design. The map, the closed form and the code are offered so that these numbers can be checked for a specific device before it is built.

Declarations

Acknowledgements: None. The author received no specific support for this work.

Funding The author declares that no funds, grants, or other support were received during the preparation of this manuscript.

Competing interests The author has no relevant financial or non-financial interests to disclose.

Ethics approval Not applicable. This is a computational study; no human participants, human data, human tissue or animals were involved.

Consent to participate Not applicable.

Consent for publication Not applicable.

Data availability All data generated in this study are produced by the simulation code described under Code availability, and the generated tables are provided as Online Resource 2. No external datasets were used.

Code availability The complete simulation, validation and figure code (Python) is provided as Online Resource 3 and will be deposited in a public repository with a permanent identifier on acceptance. It regenerates every number, table and figure in this paper.

Author contributions Ananya Sharma: conceptualization, methodology, software, validation, formal analysis, investigation, visualization, writing – original draft, writing – review and editing. The author read and approved the final manuscript.

References

1. Chou J, Li LE, Kulla E, Christodoulides N, Floriano PN, McDevitt JT (2012) Effects of sample delivery on analyte capture in porous bead sensors. Lab Chip 12:5249. https://doi.org/10.1039/c2lc40752c
2. Christensen LL (1997) Theoretical Analysis of Protein Concentration Determination Using Biosensor Technology under Conditions of Partial Mass Transport Limitation. Anal Biochem 249:153–164. https://doi.org/10.1006/abio.1997.2182
3. Ekins RP (1989) Multi-analyte immunoassay. J Pharm Biomed Anal 7:155–168. https://doi.org/10.1016/0731-7085(89)80079-2
4. Ekins RP (1998) Ligand assays: from electrophoresis to miniaturized microarrays. Clin Chem 44:2015–2030. https://doi.org/10.1093/clinchem/44.9.2015
5. Escobedo C, Brolo AG, Gordon R, Sinton D (2010) Flow-Through vs Flow-Over: Analysis of Transport and Binding in Nanohole Array Plasmonic Biosensors. Anal Chem 82:10015–10020. https://doi.org/10.1021/ac101654f
6. Foley JO, Mashadi-Hossein A, Fu E, Finlayson BA, Yager P (2008) Experimental and model investigation of the time-dependent 2-dimensional distribution of binding in a herringbone microchannel. Lab Chip 8:557. https://doi.org/10.1039/b713644g
7. Foote J, Eisen HN (1995) Kinetic and affinity limits on antibodies produced during immune responses. Proc Natl Acad Sci USA 92:1254–1256. https://doi.org/10.1073/pnas.92.5.1254
8. Fu E, Nelson KE, Ramsey SA, Foley JO, Helton K, Yager P (2009) Modeling of a Competitive Microfluidic Heterogeneous Immunoassay: Sensitivity of the Assay Response to Varying System Parameters. Anal Chem 81:3407–3413. https://doi.org/10.1021/ac802672v
9. Galvin CJ, Shirai K, Rahmani A, Masaya K, Shen AQ (2018) Total Capture, Convection-Limited Nanofluidic Immunoassays Exhibiting Nanoconfinement Effects. Anal Chem 90:3211–3219. https://doi.org/10.1021/acs.analchem.7b04664
10. Gervais T, Jensen KF (2006) Mass transport and surface reactions in microfluidic systems. Chem Eng Sci 61:1102–1121. https://doi.org/10.1016/j.ces.2005.06.024
11. Glaser R (1993) Antigen-Antibody Binding and Mass Transport by Convection and Diffusion to a Surface: A Two-Dimensional Computer Model of Binding and Dissociation Kinetics. Anal Biochem 213:152–161. https://doi.org/10.1006/abio.1993.1399
12. Harris CR, van der Millman KJ et al (2020) Array programming with NumPy. Nature 585:357–362. https://doi.org/10.1038/s41586-020-2649-2
13. Hofmann O, Voirin G, Niedermann P, Manz A (2002) Three-Dimensional Microfluidic Confinement for Efficient Sample Delivery to Biosensor Surfaces. Application to Immunoassays on Planar Optical Waveguides. Anal Chem 74:5243–5250. https://doi.org/10.1021/ac025777k
14. Hu G, Gao Y, Li D (2007) Modeling micropatterned antigen–antibody binding kinetics in a microfluidic chip. Biosens Bioelectron 22:1403–1409. https://doi.org/10.1016/j.bios.2006.06.017

15. Hunter JD (2007) Matplotlib: A 2D Graphics Environment. Comput Sci Eng 9:90–95. https://doi.org/10.1109/MCSE.2007.55
16. Jokić I, Jakšić O, Frantlović M, Jakšić Z, Guha K, Rao KS (2021) Temporal response of biochemical and biological sensors with bimodal surface adsorption from a finite sample. Microsyst Technol 27:2981–2987. https://doi.org/10.1007/s00542-020-05051-w
17. Khalil M, Teunissen CE, Otto M et al (2018) Neurofilaments as biomarkers in neurological disorders. Nature 14:577–589. https://doi.org/10.1038/s41582-018-0058-z
18. Kusnezow W, Syagailo YV, Goychuk I, Hoheisel JD, Wild DG (2006) Antibody microarrays: the crucial impact of mass transport on assay kinetics and sensitivity. Expert Rev Mol Diagn 6:111–124. https://doi.org/10.1586/14737159.6.1.111
19. Lynn NS, Šípová H, Adam P, Homola J (2013) Enhancement of affinity-based biosensors: effect of sensing chamber geometry on sensitivity. Lab Chip 13:1413. https://doi.org/10.1039/c2lc41184a
20. Lynn NS, Martínez-López JI, Bocková M et al (2014) Biosensing enhancement using passive mixing structures for microarray-based sensors. Biosens Bioelectron 54:506–514. https://doi.org/10.1016/j.bios.2013.11.027
21. Miller EA, Jabbour Al Maalouf Y, Sikes HD (2018) Design Principles for Enhancing Sensitivity in Paper-Based Diagnostics via Large-Volume Processing. Anal Chem 90:9472–9479. https://doi.org/10.1021/acs.analchem.8b02113
22. Myszka DG, He X, Dembo M, Morton TA, Goldstein B (1998) Extending the Range of Rate Constants Available from BIACORE: Interpreting Mass Transport-Influenced Binding Data. Biophys J 75:583–594. https://doi.org/10.1016/S0006-3495(98)77549-6
23. Nair PR, Alam MA (2006) Performance limits of nanobiosensors. Appl Phys Lett 88:233120. https://doi.org/10.1063/1.2211310
24. Palmqvist S, Janelidze S, Quiroz YT et al (2020) Discriminative Accuracy of Plasma Phospho-tau217 for Alzheimer Disease vs Other Neurodegenerative Disorders. JAMA 324:772. https://doi.org/10.1001/jama.2020.12134
25. Parsa H, Chin CD, Mongkolwisetwara P, Lee BW, Wang JJ, Sia SK (2008) Effect of volume- and time-based constraints on capture of analytes in microfluidic heterogeneous immunoassays. Lab Chip 8:2062. https://doi.org/10.1039/b813350f
26. Rissin DM, Kan CW, Campbell TG et al (2010) Single-molecule enzyme-linked immunosorbent assay detects serum proteins at subfemtomolar concentrations. Nat Biotechnol 28:595–599. https://doi.org/10.1038/nbt.1641
27. Saviranta P, Okon R, Brinker A, Warashina M, Eppinger J, Geierstanger BH (2004) Evaluating Sandwich Immunoassays in Microarray Format in Terms of the Ambient Analyte Regime. Clin Chem 50:1907–1920. https://doi.org/10.1373/clinchem.2004.037929
28. Shah RK, London AL (1978) Laminar Flow Forced Convection in Ducts. Academic, New York. https://doi.org/10.1016/C2013-0-06152-X

29. Sheehan PE, Whitman LJ (2005) Detection Limits for Nanoscale Biosensors. Nano Lett 5:803–807. https://doi.org/10.1021/nl050298x
30. Sjoelander S, Urbaniczky C (1991) Integrated fluid handling system for biomolecular interaction analysis. Anal Chem 63:2338–2345. https://doi.org/10.1021/ac00020a025
31. Squires TM, Messinger RJ, Manalis SR (2008) Making it stick: convection, reaction and diffusion in surface-based biosensors. Nat Biotechnol 26:417–426. https://doi.org/10.1038/nbt1388
32. Stroock AD, Dertinger SKW, Ajdari A, Mezić I, Stone HA, Whitesides GM (2002) Chaotic Mixer for Microchannels. Science 295:647–651. https://doi.org/10.1126/science.1066238
33. Vijayendran RA, Ligler FS, Leckband DE (1999) A Computational Reaction–Diffusion Model for the Analysis of Transport-Limited Kinetics. Anal Chem 71:5405–5412. https://doi.org/10.1021/ac990672b
34. Virtanen P, Gommers R, Oliphant TE et al (2020) SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods 17:261–272. https://doi.org/10.1038/s41592-019-0686-2
35. Young ME, Carroad PA, Bell RL (1980) Estimation of diffusion coefficients of proteins. Biotechnol Bioeng 22:947–955. https://doi.org/10.1002/bit.260220504
36. Zhao M, Wang X, Nolte D (2010) Mass-transport limitations in spot-based microarrays. Biomed Opt Express 1:983. https://doi.org/10.1364/boe.1.000983
37. Zimmermann M, Delamarche E, Wolf M, Hunziker P (2005) Modeling and Optimization of High-Sensitivity, Low-Volume Microfluidic-Based Surface Immunoassays. Biomed Microdevices 7:99–110. https://doi.org/10.1007/s10544-005-1587-y

Figures

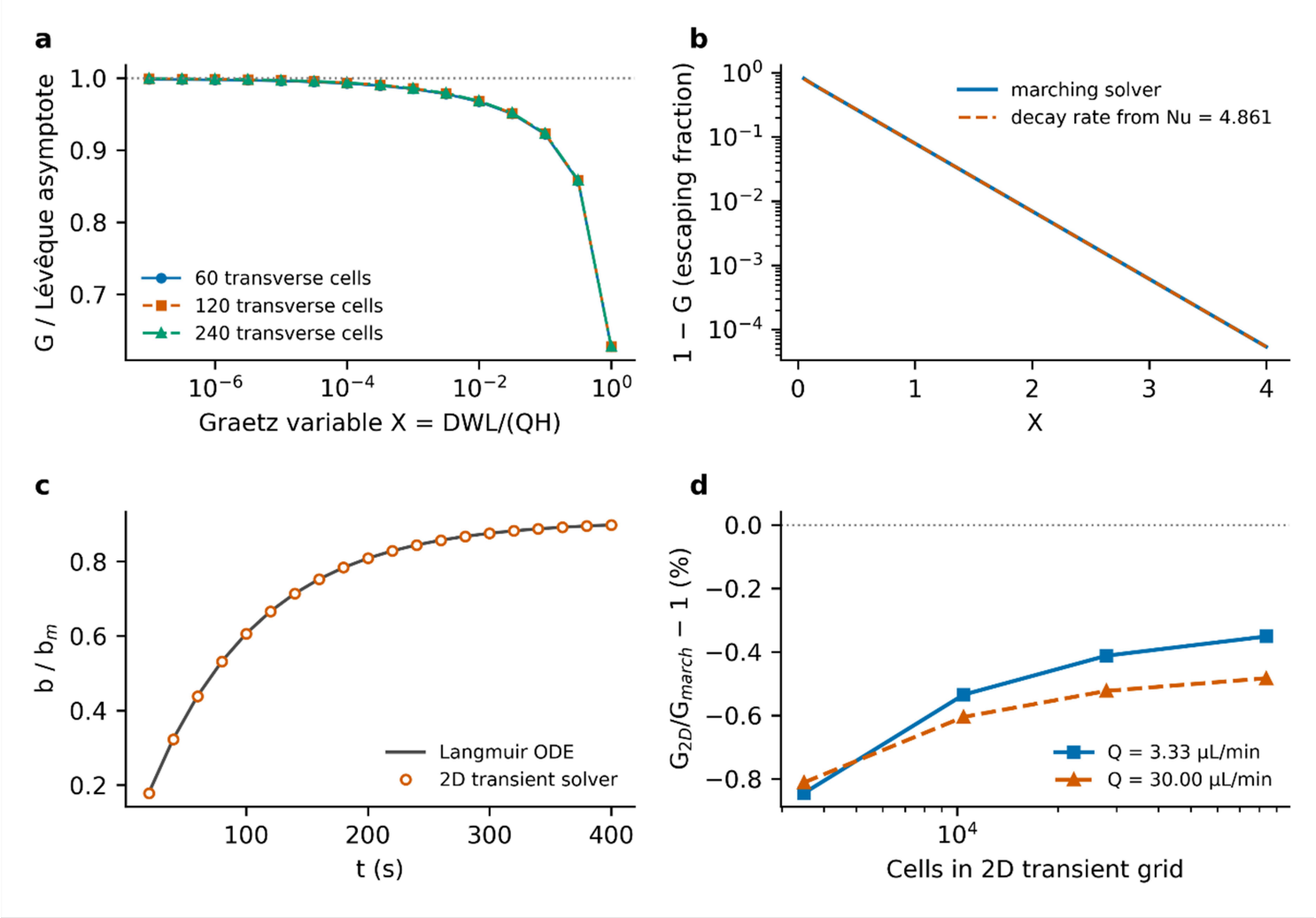


Figure 1

Validation. (a) Per-pass capture fraction G of a perfectly absorbing patch divided by the Lévêque asymptote, on three transverse grids. (b) Fraction escaping a long absorbing patch, 1 − G, compared with the fully developed decay rate implied by the tabulated Nusselt number for parallel plates with one active wall. (c) Bound density from the two-dimensional transient solver compared with the analytic Langmuir solution in a reaction-limited, saturating case. (d) Relative difference between the transient and marching solvers at steady state, at the optimum flow rate (3.33 µL min^{-1}) and at 30 µL min^{-1}, against the number of cells in the transient grid

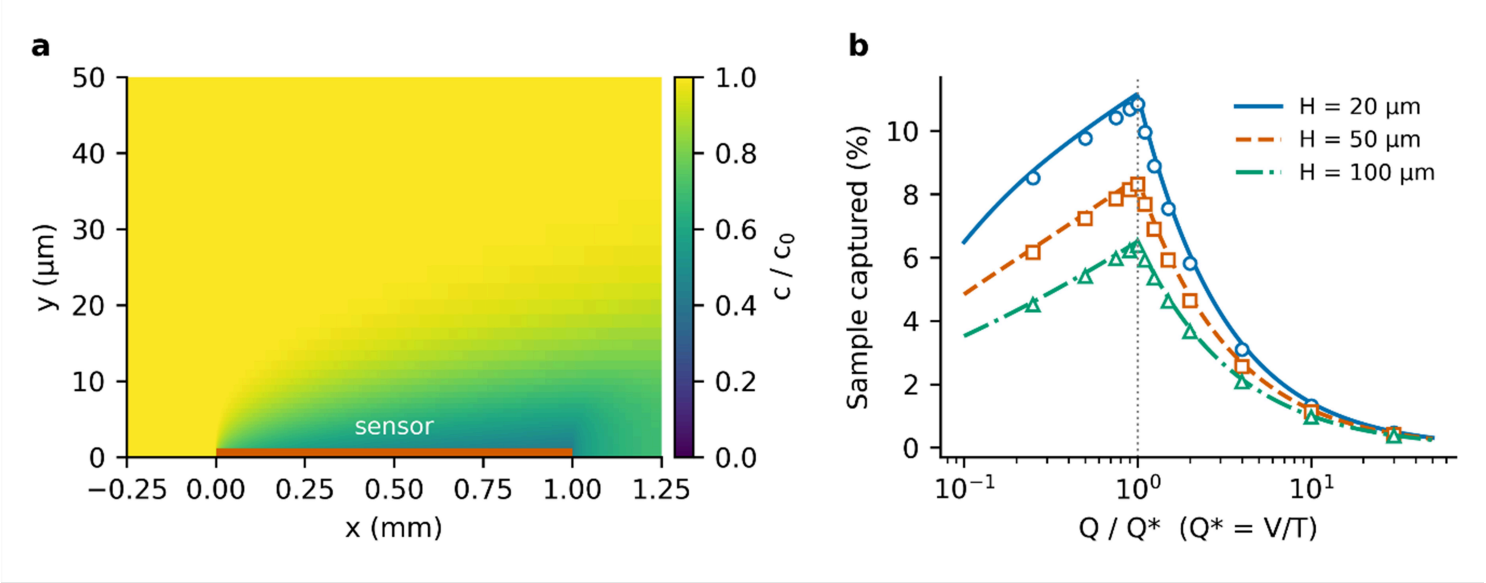


Figure 2

A finite sample has an optimum flow rate. (a) Concentration field (normalised by the inlet concentration) over the nominal capture patch after 120 s of continuous sample delivery at Q* = V/T = 3.33 µL min⁻¹, from the transient solver; the patch is marked on the floor. (b) Percentage of the sample captured at the readout time T = 15 min against flow rate scaled by Q*, for channel heights of 20, 50 and 100 µm. Lines: quasi-steady prediction, equation (3). Open symbols: two-dimensional transient simulations with the sample delivered as a plug followed by buffer. The dotted vertical line marks Q = Q*

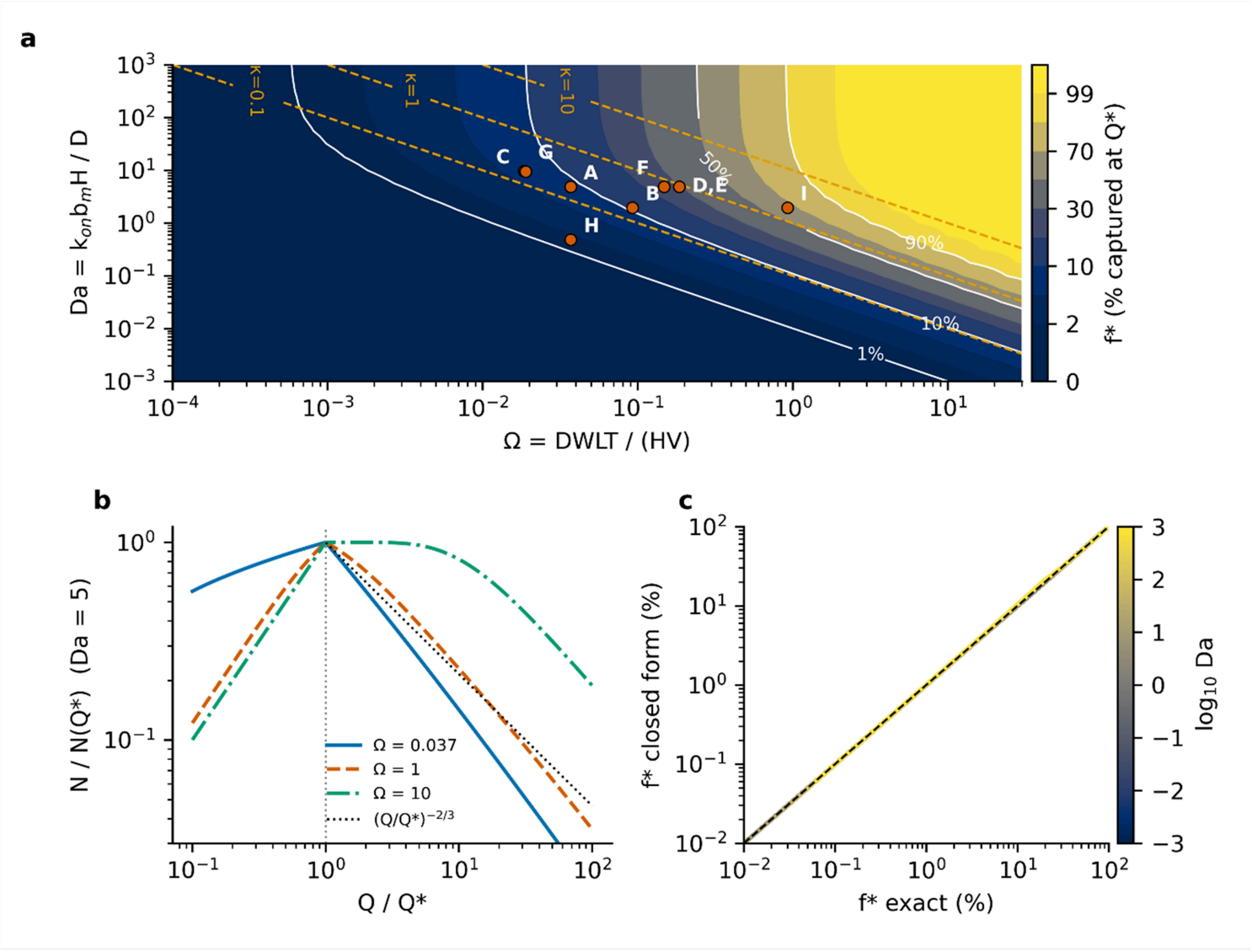


Figure 3

(a) Best achievable captured fraction f* (percentage of the sample captured at Q*) as a function of Ω = DWLT/(HV) and Da = k_{on} b_m H/D. White lines: 1%, 10%, 50% and 90% contours. Dashed lines: contours of the kinetic group κ = Ω·Da. Points: designs A–I of Table 1 (D and E coincide). (b) Captured molecules relative to the optimum against Q/Q* at Da = 5 for three values of Ω; the dotted line is the Lévêque slope $(Q/Q^*)^{-2/3}$. (c) Closed form, equation (5), against the exact f* for all 3086 tabulated points, coloured by $\log_{10}$ Da; the dashed line is equality

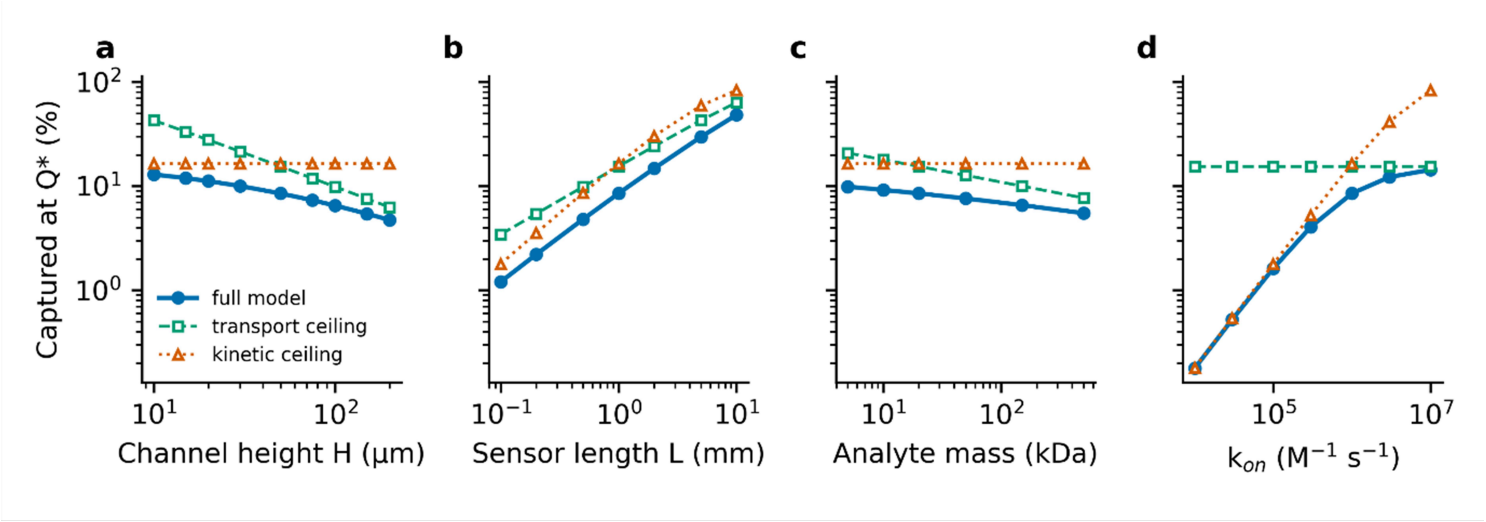


Figure 4

Effect of single design parameters on the percentage of sample captured at Q*, starting from the nominal design: (a) channel height, (b) capture-patch length, (c) analyte molecular mass (through the diffusion coefficient), (d) association rate constant. Filled circles: full model. Open squares: transport ceiling $G_{\infty}(\Omega)$ (infinitely fast binding). Open triangles: kinetic ceiling $1 - \exp(-\kappa)$ (infinitely fast transport)

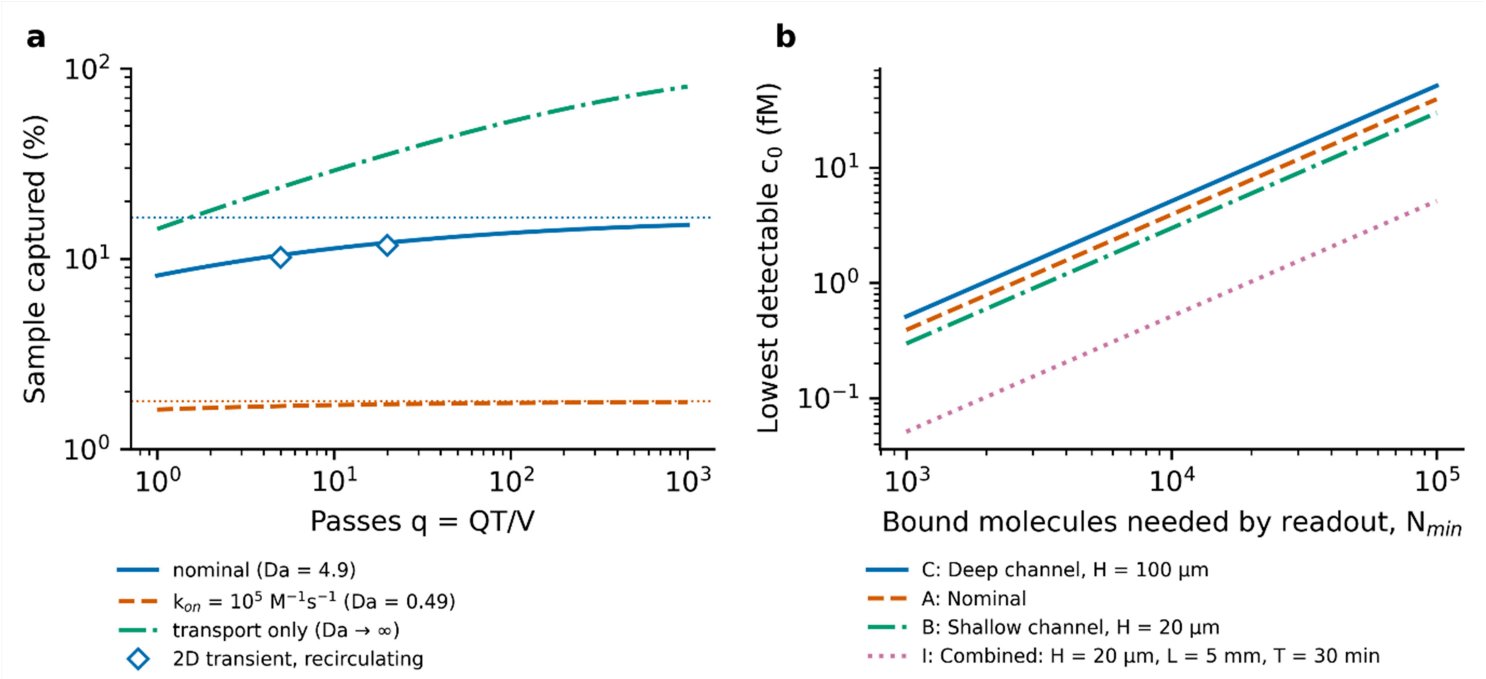


Figure 5

(a) Percentage of the nominal sample captured with recirculation, against the number of passes q = QT/V, for the nominal design, a ten-fold slower binder and the transport-only limit. Dotted horizontal lines: kinetic ceilings $1 - \exp(-\kappa)$. Open diamonds: two-dimensional transient simulations with an explicit

well-mixed recirculation reservoir. (b) Lowest detectable sample concentration against the number of bound molecules the readout needs, for four designs of Table 1 at Q*

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- OnlineResourcesinfo.docx
- ESM3.zip
- ESM1.pdf
- ESM2.zip