Solution-Path Discovery: An AI Scientist Composing Methods for Solar Cell Metallization

Technical Report v0.1, work in progress

Thinkers Lab Team

September 8, 2026

Abstract

The reported state of the art in solar cell front metallization is topology optimisation, by the method of moving asymptotes or by a learned generator, and both are reported to beat conventional finger grids. We started from those solutions and asked whether agentic discovery could push them further.

It could not, yet, and the reason is the first result. The designs held up as best carry a caveat: they are graded rather than binary, or they contain metal thinner than the process can print, and the released designs we examined are both at once. Our system applied binarity, minimum width and connectivity as hard constraints inside the optimisation loop, rejecting every design that failed one and verifying the rejection itself against cases of known answer. Under that rigour the ranking reverses, and a conventional finger grid outperforms both optimisers on less than half the metal of one of them. The system then proposed conventional families with a physical argument for each and swept them across four contact geometries, turning up strong designs that are not obvious: a symmetric comb that beats the closed-form grid, and a finger pitch the classical expression does not give.

Neither optimiser reaches the best conventional design from a structureless start on any geometry, falling short by 0.68 to 1.22 points. This is a first investigation, not a verdict. The remaining target is about 0.4 points, it must be won without adding metal, and the obstacle is the representation rather than the optimiser. We argue that a discovery harness searching over the formulation, with manufacturability enforced inside the loop, should push these methods past the conventional baselines or find solution paths neither reaches alone.

1 Introduction

Front metallization is a trade between shading and resistance. Metal carries current to the busbar and blocks the light under it, so efficiency peaks at some metal fraction in between, and the design question is where to put the metal rather than how much. The conventional answer is a grid of parallel fingers whose spacing follows in closed form from balancing emitter loss against shading. The modern answer is topology optimisation, and published comparisons report that both a gradient method and a learned generator beat the conventional grid.

We began from the reported solutions rather than from the problem. The method of moving asymptotes and the learned generator are what the literature puts forward, both are reported to beat conventional grids, and the question we set was whether agentic discovery could push them further still. We expected scope to improve them. The first thing the system found was that the comparison they win is not the one a manufacturer cares about.

The designs held up as best carry a caveat, and it takes one of two forms. Some are graded: a density field somewhere between metal and no metal, which is not a thing that can be printed. Some contain metal thinner than the process can lay down. The released designs from the work we extend are both at once, being continuous fields whose minimum feature is

a single element. The caveat is rarely stated because the constraint is rarely applied: there is no minimum-feature rule, no connectivity check and no binarisation step anywhere in that optimisation, so the reported number belongs to an object that does not exist.

Our system applied the three constraints as hard rules inside the loop rather than as checks at the end. A design must be binary, must meet a three-element minimum width, and must connect to the busbar with no isolated metal. Every design failing any of them was rejected, and the rejection was verified against cases with a known answer so that it neither passes a bad design nor discards a good one. Applying that rigour is what produced the first result, and it is not the one we set out to find: under it, a conventional design outperforms both optimisers.

The system then went after the conventional designs themselves, proposing families with a physical argument for each and sweeping them. Several of the strong ones are hard to arrive at by reasoning, and the report gives the argument alongside the number in each case.

This is a first investigation. We are not arguing that these methods cannot beat conventional designs. We think they can, and making them do so under the full constraint set is the goal this work is aimed at. Our position is that it takes a discovery harness that searches over the formulation rather than only within it, and that enforces manufacturability inside the loop rather than checking it at the end. Such a harness should be able to push these methods past the conventional baselines, or to find solution paths that neither method reaches alone, with every design printable by construction.

What the system found is set out below. The first two are the results of this report. The four that follow are observations the system made while getting there: each is measured and each is useful, but they are properties of how this problem is currently posed rather than claims that will outlive an improvement to the methods.

Result 1. The published comparison uses designs that cannot be printed. Neither method applies a minimum feature width, a binary field, or a connectivity rule. Measured from the released figures, the minimum feature of both published designs is one element, with 3.1 % and 7.3 % of their metal outside any three-element feature. We reproduce both methods to within 0.05 of their published numbers, and then requiring the designs to be printable costs up to 0.66 points and nearly triples the metal. Under that requirement a conventional finger grid beats both (Section 4).

Result 2. Strong conventional designs exist that are not obvious. The system proposed families with a physical argument for each and swept them. Centre-rooted trees place second and third on a cell whose contact runs along a whole edge, where routing is already free and nothing recommends that shape. A symmetric comb beats the closed-form grid on the 6 cm edge, so the classical construction has the shape right and the proportions wrong. And the best full-edge design we found is a finger grid at a pitch the closed-form expression does not give (Section 5).

Observations along the way.. Four more, reported because they were load-bearing for the results above. *The contact decides which family wins*: no family wins on all four geometries, and what selects between them is the one choice the published formulation holds fixed. *Binarisation needs no projection*: penalising conduction as ρ^3 and photocurrent as $(1-\rho)^3$ leaves intermediate density worse than either end, and designs reach a greyness of 0.008 unaided. *The learned generator is governed by its starting density*, which the reference states before a sigmoid so that a nominal 0.05 begins at half metal; nothing across eight network and optimiser variants moved the result by more than 0.002. *Metal is the scarce resource, not conduction*: gradient descent achieves the lowest resistive loss of any design here and still finishes last, having bought it with more than twice the metal.

Neither optimiser reaches the best conventional design from a structureless start on any geometry, falling short by 0.68 to 1.22 points. This is a first investigation, not a verdict. The remaining target is about 0.4 points, it must be won without adding metal, and the obstacle is

the representation rather than the optimiser. We argue that a discovery harness searching over the formulation, with manufacturability enforced inside the loop, should push these methods past the conventional baselines or find solution paths neither reaches alone.

2 Introduction

Front metallization is a trade between shading and resistance. Metal carries current to the busbar and blocks the light under it, so efficiency peaks at some metal fraction in between, and the design question is where to put the metal rather than how much. The conventional answer is a grid of parallel fingers whose spacing follows in closed form from balancing emitter loss against shading. The modern answer is topology optimisation, and published comparisons report that both a gradient method and a learned generator beat the conventional grid.

We began from the reported solutions rather than from the problem. The method of moving asymptotes and the learned generator are what the literature puts forward, both are reported to beat conventional grids, and the question we set was whether agentic discovery could push them further still. We expected scope to improve them. The first thing the system found was that the comparison they win is not the one a manufacturer cares about.

The designs held up as best carry a caveat, and it takes one of two forms. Some are graded: a density field somewhere between metal and no metal, which is not a thing that can be printed. Some contain metal thinner than the process can lay down. The released designs from the work we extend are both at once, being continuous fields whose minimum feature is a single element. The caveat is rarely stated because the constraint is rarely applied: there is no minimum-feature rule, no connectivity check and no binarisation step anywhere in that optimisation, so the reported number belongs to an object that does not exist.

Our system applied the three constraints as hard rules inside the loop rather than as checks at the end. A design must be binary, must meet a three-element minimum width, and must connect to the busbar with no isolated metal. Every design failing any of them was rejected, and the rejection was verified against cases with a known answer so that it neither passes a bad design nor discards a good one. Applying that rigour is what produced the first result, and it is not the one we set out to find: under it, a conventional design outperforms both optimisers.

The system then went after the conventional designs themselves, proposing families with a physical argument for each and sweeping them. Several of the strong ones are hard to arrive at by reasoning, and the report gives the argument alongside the number in each case.

This is a first investigation. We are not arguing that these methods cannot beat conventional designs. We think they can, and making them do so under the full constraint set is the goal this work is aimed at. Our position is that it takes a discovery harness that searches over the formulation rather than only within it, and that enforces manufacturability inside the loop rather than checking it at the end. Such a harness should be able to push these methods past the conventional baselines, or to find solution paths that neither method reaches alone, with every design printable by construction.

What the system found is set out below. Three of the six were not what we expected, and two of them contradict the published account of this problem.

1. The published comparison uses designs that cannot be printed. Neither method applies a minimum feature width, a binary field, or a connectivity rule. Measured from the released figures, the minimum feature of both published designs is one element, with 3.1 % and 7.3 % of their metal outside any three-element feature. We reproduce both methods to within 0.05 of their published numbers, and then requiring the designs to be printable costs up to 0.66 points and nearly triples the metal. Under that requirement a conventional finger grid beats both (Section 4).

2. The contact decides which design family wins. Sweeping a set of families across four contact geometries, no family wins everywhere. The finger grid takes the whole edge and

nothing at a corner; the comb takes the centre; the corner-rooted tree loses on both other geometries and takes the corner. What selects between them is the contact, which is the one choice the published formulation holds fixed (Section 5).

3. Some strong conventional designs have no argument behind them. Centre-rooted trees place second and third on a cell whose contact runs along a whole edge, where routing is already free and nothing recommends that shape. A symmetric comb beats the closed-form grid on the 6 cm edge, so the classical construction has the shape right and the proportions wrong. And the best full-edge design we found is a finger grid at a pitch the closed-form expression does not give (Section 5).

4. Binarisation needs no projection. Penalising conduction as ρ^3 and photocurrent as $(1-\rho)^3$ makes a half-density element nearly useless as a conductor and nearly opaque at once, so it is worse than both ends and the optimiser empties it. Held at an exponent of 3 from the first step, designs reach a greyness of 0.008 with no Heaviside projection at all (Section 6).

5. The starting density decides the learned generator, and it is easy to misread. The reference implementation sets a variable that is then passed through a sigmoid, so a stated density of 0.05 means the run begins at 0.5125: at half metal, removing material rather than adding it. On the published cell the generator returns 3.80 from a 0.20 start and 12.49 from 0.5125, with everything else identical. Nothing in the network or the training recipe moves the number by more than 0.002 (Section 6).

6. Metal is the scarce resource, not conduction. Of the 1.24 points between the conventional grid and a perfect metal that is transparent and has no resistance, only about 0.4 is resistive loss that routing could recover. Gradient descent shows what ignoring this costs: on the published cell it achieves the lowest resistive loss of any design in this work and still finishes last, because it bought that conduction with more than twice the metal (Section 5).

Neither optimiser reaches the best conventional design from a structureless start on any of the four geometries, falling short by 0.68 to 1.22 points. The learned generator leads the gradient method on all four, which is not the order the literature implies, and does so on half the metal.

We report the closing result as negative, and the target as measured rather than guessed. Of the 1.24 points between the conventional grid and a perfect metal that is transparent and has no resistance, only about 0.4 is resistive loss that better routing could recover, and it has to be taken without adding metal. That is the target. The published comparison never had to state it because it never charged for buildability. There is much more to do here. What is not yet true is that these methods, as they stand, are the way to do it.

3 Problem and model

We use the model of Gupta et al. (2015) as implemented by Bhattacharya et al. (2021). A design is a density field $\rho \in [0,1]^{n_y \times n_x}$ on a quadrilateral grid over one face of a square cell, with $\rho = 1$ metal and $\rho = 0$ bare emitter. Sheet conduction on the front surface gives

$$G(\rho)\, V = I(V), \qquad \sigma(\rho) = \sigma_{\min} + \rho^{\,p}\,(\sigma_{\text{metal}} - \sigma_{\min})\,, \tag{1}$$

with G assembled from four-node bilinear elements. The local current density couples a photocurrent, scaled by the non-metal fraction, to a one-diode dark current:

$$j = j_L\,(1-\rho)^{p_s} \;-\; j_0\left(e^{\beta V} - 1\right). \tag{2}$$

Output power is taken at the busbar and efficiency follows (Bhattacharya et al., 2021): $P_{\text{out}} = V_{\text{bus}} \sum I$ and $\eta = (P_{\text{out}}/A_c)/P_{\text{in}} \times 100$ with $P_{\text{in}} = 1000\,\text{W m}^{-2}$. Equation (1) is nonlinear through (2) and is solved by Newton–Raphson. Constants are in Appendix E.

The two exponents are not the same kind of object.. The exponent p_s on the light term is nonlinear for a physical reason: an intermediate ρ is not "half the area covered", it is partial metal, and partial metal does not transmit light in proportion to how much of it there is. Its steepness is also deliberate. It keeps the optimizer out of a semi-transparent regime that is not printable and whose optical and contact behaviour this model is not entitled to interpolate. The exponent p on the conduction term has no such defence: a partly metallized sheet plausibly conducts roughly in proportion to its metal, so $p = 1$ is arguable. Measured, it is not a free choice.

How much is on the table.. Before comparing methods it is worth bounding what any of them can achieve. A perfect collector, no shading, no resistive drop, the whole cell at V_{bus}, leaving only the diode loss, gives $j = j_L + j_0(e^{\beta V_{\mathrm{bus}}} - 1) = 288.16\,\mathrm{A\,m^{-2}}$ and hence 14.41 %. Against that ceiling the best two-variable grid reaches 12.51 % and the best result in this paper 13.20 %, so **1.20 percentage points remain**. Every number below should be read against that margin: an improvement of 0.3 points is a quarter of what is left, and a method that promises more than 1.2 is promising something impossible.

Regularization.. Density-based TO requires a length scale. We use a cone filter of radius R elements followed by a Heaviside projection (Wang et al., 2011)

$$\bar{\rho} = \frac{\tanh(\beta_p \eta_p) + \tanh\big(\beta_p(\tilde{\rho} - \eta_p)\big)}{\tanh(\beta_p \eta_p) + \tanh\big(\beta_p(1 - \eta_p)\big)}, \qquad \eta_p = 0.5, \tag{3}$$

with β_p ramped during optimization.

4 The constraint that decides the comparison

A metallization design has to be built. That puts three limits on it. Every metal feature must be at least as wide as the printer can lay down. The design must be metal or not metal, because a density of one half cannot be printed. And every finger must connect to the busbar, because metal that connects to nothing carries no current but still blocks light.

We apply all three. The minimum feature is three elements, checked by a morphological opening. The field must be binary. Every piece of metal must connect to the contact, with no isolated islands and no diagonal-only bridges. At the grid and cell size used here, three elements is 225 µm.

The first finding is that the work we set out to extend does not apply these limits, and that applying them reverses its result. The caveat in the published designs takes both forms at once. They are graded rather than binary, since the optimisation contains no binarisation step and the reported number belongs to a continuous density field. And they contain metal thinner than the process can print. Neither is stated, because neither constraint is applied.

Our system applies all three as hard rules inside the loop and rejects any design that fails one. The rejection was itself checked: a design known to meet the width rule passes it, and a rail lying against the cell boundary is eroded from outside rather than credited with material it does not have, so a too-narrow edge feature fails as it should. That matters here, because the result below depends on rejections being right rather than merely strict. Figure 1 shows the designs this section compares.

What the published designs satisfy.. We measured the released figures of the reference implementation. The procedure was checked first on a design already known to pass, so we know it does not create failures that are not there. The minimum feature of both published designs is one element. Most of each design is three to four elements wide, so the problem

Table 1: The published regime, 1.5 cm cell. As drawn, both methods reproduce the reference numbers and beat the conventional design. Once the designs must be printable, both fall below it.

	as drawn	printable	metal (%)
reference, gradient descent	12.9178	not enforced	n/a
reference, learned generator	12.9217	not enforced	n/a
this work, gradient descent	12.9645	12.3035	12.24
this work, learned generator	12.6257	12.4868	6.51
conventional finger grid	n/a	**13.1686**	**5.43**

is not the whole design. It is the branch tips and the ends of fingers. Even so, 3.1 % of the metal in the published network design and 7.3 % in the published gradient design sits outside any three-element feature and would not print. Neither published number is a printable one. The optimisation does not apply the limits either: there is no minimum-feature rule, no connectivity check and no binarisation step anywhere in it, so the reported number belongs to a continuous density field.

What that costs.. We reproduced the reference setup with the same cell, grid, filter radius and physical constants. As drawn, gradient descent gives 12.9645 against the published 12.9178, and the network gives 12.6257. Both methods reproduce. The number does not survive the constraints. Repairing the same designs so that every element sits inside a three-element block of metal costs gradient descent 0.66 points and raises its metal from 4.69 % to 12.24 %.

Table 1 contains the whole result. Read the first column and the optimisers win. Read the second, which is the only column a factory can use, and a conventional finger grid beats both, on less than half the metal of one of them. This is not caused by a harsh repair. The conventional design meets the same rule with no metal outside a three-element feature, and every number in the table comes from the same code.

Why this matters.. The published comparison is between optimisers, and on its own terms it is right. But it compares designs that cannot be built, and its ranking does not hold for designs that can. An engineer following it would pick a method that, once the printing limit is applied, loses to a design you can write down in closed form from finger-pitch theory.

5 Conventional designs, and why they are not obvious

The printable optimum turned out to be a conventional design, so the next question was how good conventional designs can be. The system proposed the families, gave a physical argument for each, and swept them. The second finding is that the strong ones are not the ones you would guess, and that several are hard to arrive at by reasoning. We give the argument beside the result in each case, because a family that wins without one is a coincidence until it is explained.

Why these families.. Each family was included because some physical argument says it should win somewhere. The sweep decides where.

The uniform finger grid is the closed-form answer when the busbar covers a whole edge. Every point of the cell then has a direct path to the contact through the emitter, so routing is free. All that is left is to keep the average distance from any point to metal small at a fixed metal area, and parallel equally spaced lines do that exactly. The best pitch follows from balancing emitter loss against shading.

The tapered rail follows from the current a member carries. A collector running away from the contact carries less current the further it goes, because it has already delivered what it picked up. A member of constant width is therefore too wide at its far end, and pays shading for conduction it does not need. Tapering it to the minimum width at the tip is worth 0.0840 on the centre contact, at lower metal.

The corner-rooted tree is the Murray-law construction. When the contact is a point rather than an edge, current has to be routed before it can be collected. A branching structure whose members narrow by a fixed exponent at each split is the classical answer for routing at least loss.

The symmetric comb exists because a contact has a symmetry and the best design usually has the same symmetry. We added this family after the sweep returned only lopsided designs on a symmetric contact, which was a gap in the candidate set rather than a failure of the search. Adding it was worth 0.6973, more than any optimiser gained on that geometry.

Rings, radial fans and crosses were included to be beaten. They are what you draw when asked for something that is not a grid, and their job is to show that the families with a physical argument behind them do better.

What the sweep found.. No family wins everywhere, and what decides between them is the contact. The finger grid takes the whole edge and nothing at a corner. The comb, which contains the finger grid as its one-spine case, takes the centre. The corner-rooted tree loses on both other geometries and takes the corner. So the thing that picks the family is the contact, which is the one choice the published work holds fixed.

Three results were not obvious. On the 1.5 cm cell the second and third best designs are centre-rooted trees, on a cell whose contact runs along a whole edge. Nothing recommends that shape there, since routing is already free, yet both land within 0.5 points of the closed-form answer. On the 6 cm whole edge a symmetric comb beats the plain grid by 0.0496, so the classical construction has the shape right and the proportions wrong. And the best full-edge design we found anywhere is a finger grid at a pitch the closed-form expression does not give: 10.9458 on 15 % metal against the classical grid's 10.7914 on 18 %. More efficiency on a fifth less silver.

Appendix D shows the best designs on each contact and Appendix C describes the three 6 cm geometries. Appendix I lists every family swept.

How far conventional designs are from the ceiling, and what is scarce.. If every node sat at the busbar voltage and no metal blocked any light, the cell would give 14.4080 %. With no metal at all it gives 3.8047 %. So the metal has 10.6033 points to win, and the conventional grid takes 88.3 % of them. We split what is left into two parts: the light the design has already given up to its own metal, and the resistive loss that better routing could still recover. Only 0.3978 points of the grid's remainder is resistive loss. So the target for any search is not the 1.24 points between the grid and the ceiling. It is about 0.4, and it has to be taken without adding metal, because each extra 1 % of metal costs about 0.15 points of shading.

That reverses the usual intuition about this problem. Conduction is not the scarce resource. Metal is, and a method that spends metal to buy conduction is trading the wrong way.

6 What was tried on the two methods

The conventional designs set a bar. The rest of the budget went on the question of whether either optimiser reaches it once the printing limits are applied. This section says what was varied and where it ended up.

Table 2: Binarisation schedules, printable efficiency, 1.5 cm cell. Three give the same answer and the fourth gives nothing printable. On the objective each optimiser reports for itself, these four differ by more than eight points.

schedule	gradient descent	learned generator
exponent 3, no projection	12.3035	12.4868
exponent 3, then projection	12.2745	12.4982
exponent ramped, no projection	12.4881	12.4982
exponent ramped, then projection	none	none

Starting density.. This mattered more than anything else we varied, and it is easy to get wrong. The reference implementation sets a variable that is then passed through a sigmoid, so a stated density of 0.05 is a value before that function is applied, and the run actually starts at 0.5125. It starts at half metal and takes material away rather than adding it. We swept starting densities of 0.02, 0.055, 0.09, 0.20 and 0.5125, plus an untrained random generator. The spread is wide. On the published cell the network gives 3.80 from a 0.20 start and 12.49 from 0.5125, with everything else the same. Gradient descent is much less sensitive and varies by under a point across the same range.

Length scale.. The density filter is what sets a minimum feature size during the run. A rule checked only at the end is a test, not a constraint. We ran filter radii of 1.5, 2, 3, 4, 5, 6, 8, 10 and 12 elements. The response is not monotone. Too small and the design uses features finer than the printer can make. Too large and the filter smooths away the fine grid the problem wants. All results here use a radius of 1.5 elements, which is what the reference implementation uses.

Binarisation, and why no projection is needed.. Grey material is removed by the SIMP exponents, not by a projection. This is a result rather than a setting: the usual Heaviside continuation is unnecessary on this problem. Conduction scales as ρ^3 and photocurrent as $(1-\rho)^3$, so an element at half density is nearly useless as a conductor and nearly opaque at the same time. It is worse than both ends, so the optimiser empties it. With the exponent held at 3 from the first step, designs go binary on their own, to a greyness of 0.008, with no projection at all. We compared four schedules: the exponent fixed or ramped, each with and without a projection.

Table 2 carries a warning that is worth more than the schedule itself. Ranked on the objective each optimiser reports for itself, the four differ by over eight points and a clear winner appears. Ranked on whether the design can be printed, three are the same within 0.21 and 0.011, and the fourth gives nothing. A comparison made on the wrong quantity would have looked confident and been wrong.

Network and optimiser.. We ran eight configurations to a budget of 8976 evaluations at a fixed start and filter radius: kernel size 3 against 5; per-pixel offset scale 10, 25 and 50; a wider network with 256 filters and latent size 256; the upsampling moved earlier so more convolutions run at full resolution; AdamW with weight decay and a cosine schedule against plain Adam; and an attention generator in place of the convolutional one. All eight landed within 0.002 of each other as drawn. Nothing in the network or the training recipe changed the answer. The starting density and the filter radius did.

Budget.. Budgets ran from 40 to 1500 evaluations per arm, with ladders of up to eight rungs. Budget matters when the schedule is wrong and not otherwise. Gradient descent gains 4.6

Table 3: Printable efficiency from a start with no structure, against the best conventional design on each geometry. Every design here is binary, meets the three-element minimum, and has no floating metal.

cell	contact	best conventional	learned generator	gradient descent
1.5 cm	whole edge	**13.1686**	12.4868	12.3035
6 cm	whole edge	**10.9458**	9.7194	7.8733
6 cm	centre pad	**10.5792**	9.9069	9.4680
6 cm	corner	**9.4579**	8.7550	7.8947

points going from 40 to 150 evaluations under a poor schedule, and gains nothing past a few hundred once the schedule is right.

Appendix G gives the searcher settings, and Appendix B describes how the hypotheses behind these runs were generated.

Where the methods land.. Table 3 gives the result under the full set of limits, on four contact geometries, starting from no structure.

Neither optimiser reaches the conventional design on any of the four geometries. The gap runs from 0.68 to 1.22 points. The network is ahead of gradient descent on all four, which is not the order the published comparison implies, and it does so on much less metal: 6.51 % against 12.24 % on the published cell.

Gradient descent fails in a way worth describing. On that cell it has the lowest resistive loss of any design in this work, 0.2077 against the conventional grid's 0.3978, and still comes last. It bought that conduction with 12.24 % metal and gave back 1.8968 in shading. Conduction is not what is scarce on this problem. Metal is.

Why we expect this to be recoverable.. Two things in this table argue against reading it as a limit of the methods. The learned generator reaches its number on 6.51 % metal where gradient descent needs 12.24 %, so it is already finding efficient structure rather than spending silver to hide a poor one. And nothing in the network or the training recipe moved the result by more than 0.002 across eight configurations, which places the ceiling outside the architecture and outside the optimiser. What remains is the representation. The minimum feature is currently imposed by a filter and checked afterwards, so a design that is good before printing need not survive it. A parameterisation that carries the length scale structurally would remove that failure mode rather than tune around it, and that is the direction this work is aimed at.

7 What this does not establish

The model is a single-diode formulation on a uniform mesh with fixed material constants and a fixed busbar voltage, and no result here has been validated against a fabricated cell. Absolute efficiencies should be read to one decimal: a 1 % change in the diode parameter moves them by 0.18, which is more than several of the differences discussed. Comparisons between designs are made at identical parameters, where that uncertainty is common-mode, and only orderings that survive 10 % perturbations of the diode parameter, the photocurrent and the metal conductance are reported as orderings.

No result here has been validated against a fabricated cell, and the model's own accuracy against measurement is not established by this work. The scoring, selection and plotting code is released and every figure regenerates from the design arrays with it, but the generator for the parameterised family candidates is not included at this revision, so those designs are distributed as arrays rather than as code.

The reference designs are measured from released figures rather than from released arrays, so their feature statistics carry a resampling uncertainty of about one element; the procedure was verified not to corrupt a design known to be compliant. The negative result on the two methods is a statement about the configurations searched, which are listed in Section 6 and the appendices, and not a proof that no configuration exists. The autonomous system generated and screened the hypotheses reported here; the experiments it ran are recorded, but its own selection among candidate hypotheses is not itself evaluated in this revision.

8 Related work

Metallization design.. Gupta et al. (2015) optimise front metallization patterns by topology optimisation, and Bhattacharya et al. (2021) apply a convolutional generator to the same problem and report that it and the method of moving asymptotes both beat conventional grids. This report re-examines that comparison. We reproduce both methods to within 0.05 of their reported values and find that the comparison is made on designs that carry metal one element wide, so its ranking does not hold once the designs must be printable.

Length scale and printability.. Minimum length scale, mesh dependence and the robust formulation are long established in structural topology optimisation. Sigmund (2007) sets out morphology-based filters, Wang et al. (2011) the robust formulation over eroded, nominal and dilated designs, and Zhou et al. (2015) minimum length scale through geometric constraints. We claim none of this machinery as new. What we add is its consequence for this problem: a conduction and shading objective with two penalisation exponents, where applying the length scale changes which method wins rather than only what the winner looks like.

Parameterisation.. Hoyer et al. (2019) show that optimising the weights of a generator rather than a density field improves structural optimisation, which is the method Bhattacharya et al. (2021) carry into metallization. Our results are consistent with that on the objective those papers report, and reverse on a printable one. We also find the generator far more sensitive to its starting density than direct gradient descent, which matters because the starting density is rarely stated. Svanberg (1987) is the gradient method used throughout as the comparison arm.

Automated search over method choices.. Searching over the formulation rather than within it resembles automated machine learning pipeline search, and we make no claim of novelty for the idea. The difference here is what is searched: the exponents, the schedule, the filter radius, the starting density and the contact, which are choices usually made once and not recorded. Our finding that these are worth more than the choice of optimiser is the reason we think that search is worth automating.

9 Conclusion

We started from the reported solutions and asked whether agentic discovery could push the method of moving asymptotes and the learned generator past conventional designs. Two results came back.

The comparison those methods win is made on designs that cannot be printed. Applying binarity, minimum width and connectivity as hard constraints inside the loop, and rejecting every design that fails one, reverses the ranking: a conventional finger grid outperforms both. And the conventional space itself holds strong designs that are not obvious, which the system proposed with a physical argument for each and which a sweep confirmed.

Four further observations came out along the way, and we report them because they were load-bearing rather than because we expect them to last: the contact decides which family

wins, binarisation needs no projection, the learned generator is governed by a starting density that is easy to misread, and metal rather than conduction is what is scarce. If these methods improve, most of those become properties of a formulation that has been superseded.

Under the full constraint set neither method reaches the conventional baseline on any of four geometries. We read that as the outcome of a first investigation rather than a property of the methods, and the measurements say why. The shortfall is 0.68 points at its smallest, against a target of about 0.4 that is all any method can win by routing. The learned generator gets there on 6.51 % metal where gradient descent needs 12.24 %, so it is finding efficient structure rather than spending silver. Nothing in the network or the training recipe moves the result by more than 0.002, which puts the ceiling outside the architecture and outside the optimiser. And the failure has a location: the length scale is imposed by a filter and checked afterwards, so a design that is good before printing need not survive it.

That is what the next step has to address. A parameterisation emitting centrelines and widths would make manufacturability a property of the representation rather than a test applied at the end. Beyond it, the formulation is worth searching: the exponents, the schedules, the filter and the contact, each of which mattered more here than the choice of optimiser. A discovery harness over those choices, with manufacturability enforced inside the loop, is what we think it will take to push these methods past the conventional baselines, or to find solution paths neither reaches alone. That expectation is not established here. What is established is that the target is real, that it is about 0.4 points, that it must be taken without adding metal, and that the obstacle is the representation. The loop that produced these results is not yet closed, and Section B.3 says where it is open.

Acknowledgement

This report re-examines the implementation released by Bhattacharya et al. (2021). The comparison in Section 4 was only possible because that code and its designs are public.

References

Sumit Bhattacharya, Devanshu Arya, Debjani Bhowmick, Rajat Mani Thomas, and Deepak K. Gupta. Improving solar cell metallization designs using convolutional neural networks. *arXiv preprint arXiv:2104.04017*, 2021.

Deepak K. Gupta, Matthijs Langelaar, Marco Barink, and Fred van Keulen. Topology optimization of front metallization patterns for solar cells. *Structural and Multidisciplinary Optimization*, 51(4):941–955, 2015.

Stephan Hoyer, Jascha Sohl-Dickstein, and Sam Greydanus. Neural reparameterization improves structural optimization. In *NeurIPS Workshop on Solving Inverse Problems with Deep Networks*, 2019. arXiv:1909.04240.

Ole Sigmund. Morphology-based black and white filters for topology optimization. *Structural and Multidisciplinary Optimization*, 33(4–5):401–424, 2007.

Krister Svanberg. The method of moving asymptotes, a new method for structural optimization. *International Journal for Numerical Methods in Engineering*, 24(2):359–373, 1987.

Fengwen Wang, Boyan S. Lazarov, and Ole Sigmund. On projection methods, convergence and robust formulations in topology optimization. *Structural and Multidisciplinary Optimization*, 43(6):767–784, 2011.

Table 4: Projection continuation, MMA, 40×40. As β_p rises the optimized objective and the printed design converge, and the printed design improves by nearly three points.

β_p	optimized η (%)	printed η (%)	gap	grey fraction
0	11.76	9.20	2.56	0.22
4	12.24	10.88	1.36	0.21
8	12.54	11.38	1.16	0.11
32	11.94	12.07	−0.14	0.02
64	11.99	12.07	−0.08	0.01

Mingdong Zhou, Boyan S. Lazarov, Fengwen Wang, and Ole Sigmund. Minimum length scale in topology optimization by geometric constraints. *Computer Methods in Applied Mechanics and Engineering*, 293:266–282, 2015.

A Grey material and what is actually printed

The filter that imposes a length scale also creates intermediate density, and the exponents of Section 3 penalize it, but nothing removes it. At 40×40 with $R = 1.5$ a one-element finger filters to a peak of $\bar{\rho} = 0.79$: it pays $0.79^3 = 0.49$ of the metal conductance while blocking essentially all light beneath it. Nothing in the design reaches $\rho = 1$.

We therefore evaluate, alongside the optimized objective, the design as it would be *printed*: the projected field thresholded at η_p.

Table 4 follows one run up the projection ladder.

Ending at high β_p is a requirement, not a tuning preference: a run that stops while grey remains has optimized a cell that cannot be made. Thresholding at the end is not a substitute. It measures how much the optimizer was misled, and on these designs it costs 1.5 points. The design must become binary *during* the search.

B An AI scientist searching the space

Earlier work in this programme established that solution-path choice dominates method choice by two orders of magnitude, and that the winning path was found by a short sequence of hypotheses rather than by enumeration. We now ask whether that sequence can be produced by a language model.

The loop has three stages: propose hypotheses over solution paths; screen them cheaply and discard most; test the survivors and re-hypothesize from what survives. **The first stage has been run. The second and third have not**, and Section 7 states that plainly.

B.1 What was run: hypothesis generation

Two rounds of generation were carried out, each with several independent agents given the problem statement, the physics, and the measurements of the facts established earlier in this programme. Agents were told to assume any evaluator could be built, so that proposals were not truncated to fit existing tooling; the cost of building each was assessed afterwards. One round applied adversarial screening; the other deliberately did not, so that an unscreened batch could be measured.

Every proposal is required to carry: a mechanism; a prediction with a margin specific enough to be wrong; a falsifier; **the axis its mechanism depends on**; and a screening test that is cheap along a *different* axis.

Table 5: Two rounds of agent-generated hypotheses over solution paths. "Entangled" counts proposals whose own screening test is cheap along the axis their mechanism depends on, and which would therefore reject themselves.

round	proposals	entangled screens	thin mechanisms
Breadth (9 angles, unscreened)	81	22 (27%)	0
Unconfined (5 angles, screened)	30	8 (27%)	0

Why that last requirement exists.. A screen cheap along the axis a mechanism depends on rejects hypotheses that are true. This is not hypothetical. The composition that produces this paper's best result screens at 40×40 as a 0.012-point gain, indistinguishable from noise, and is 0.694 at 160×160, because its mechanism *is* resolution. Screening it cheaply on a coarse mesh would have killed it. The same error was then made a second time, on a different axis, within the same working session and after the rule had been written down.[1] A rule broken twice by its author within an hour is not a rule, so it is checked mechanically by a validator over the register.

B.2 What the generation produced

Table 5 lists what the two rounds produced.

Two results, in opposite directions.

Mechanisms were sound.. Not one of the 111 proposals was a bare method name. Every one supplied a reason specific to this problem, most referring to a measured fact and predicting a consequence from it. The sharpest identified a mechanism the authors had missed: that the optical loss is a pure area integral, exactly linear in metal area, so on a pixel field with a one-element minimum feature it is *quantized* in steps of one element area, a first-order discontinuity in the objective at precisely the scale the design cares about, which is a candidate explanation for the mesh-tracking of the mesh-convergence study reported separately.

Screens were entangled at a consistent rate.. 27% of proposals in each round, flagged independently, by an adversarial screening agent in one round and by the mechanical validator in the other, proposed a test that would reject their own hypothesis. The agreement across two independent detection methods and two separate batches makes this a property of the generation rather than an artefact. It is also not obvious on reading: several entangled screens are subtle, and a human reviewer would plausibly approve most of them.

The register.. All 126 hypotheses, their mechanisms, predictions, falsifiers, screening axes and verdicts are in the hypothesis register released with this report, generated directly from the machine-readable register so that paper and register cannot drift. Five are validated, two refuted, and the history of both refutations records that one was initially mis-scoped and corrected, because a register that hides its own errors is a worse instrument than one that does not.

B.3 What was not run

This report should not be read as describing a closed loop, and the gap is worth stating precisely.

[1] A hypothesis that the fixed terminal voltage misranks designs was screened by sweeping finger pitch at one busbar geometry, and read as refuted. The mechanism depends on series resistance, which is set by the contact geometry rather than by the pitch; varying that instead shows the effect reaches 0.78 points.

No language model proposed a formulation directly to the optimisation engine, and no experiment reported here was selected by one without a person in the loop. Hypotheses were generated and screened, but the screening was by argument rather than against the evaluator, and it ran on one batch. Nothing was re-hypothesised from a screened result inside the loop itself.

Three things would close it. Screening should run against the evaluator rather than by argument, so a hypothesis is refuted by measurement. An expert ranking should be committed to version control before any autonomous run, so that comparing the system against a person is a measurement rather than a recollection. And the system should report a calibration rate, meaning the fraction of its stated predictions that turn out correct, which is a property of the reasoning rather than of the design it produced. We consider that last number the most informative one, and this revision does not have it.

C Three contact geometries, and what each one taught

The published starting point for front metallization is a parallel finger grid feeding a busbar that runs the length of one edge. That is the design the literature optimizes, and the variable it optimizes is the finger pitch. This section reports what changed when the contact itself was treated as the independent variable, across three geometries on one cell and one model.

The organisation is deliberate. Each geometry answered a question at a different level, and the levels are not interchangeable: a programme-level claim that survives one geometry is falsified by the next, and the mechanism that explains the third was invisible until the second had been measured.

C.1 Level one: does a density field earn its variables?

The premise behind a free density field is that it discovers designs a feature-based parameterization cannot express. On the full-edge busbar it does not, and the report gives the current numbers for that comparison.

The reason is geometric rather than numerical. With the busbar along a whole edge, every point already has a direct path out through the emitter, so routing is free and the only task left is to minimize mean distance to metal at fixed metal area. Parallel lines solve that exactly. Adding a vertical spine to the best finger grid costs 0.1141.

C.2 Level two: the contact decides whether the question exists

The same spine, added to the same finger grid, on a centre pad instead of an edge, gains 9.7603.

busbar	best fingers (%)	fingers + spine (%)	change
whole left edge	11.0007	10.8866	−0.1141
centre pad	0.0653	9.8256	+9.7603

This is the result that reframes the programme. Whether topology optimization is worth doing at all is not a property of the objective, the solver or the resolution. It is a property of the contact, and the contact is the one thing the published formulation holds fixed. A finger grid on a centre pad scores 0.0653 because most of its metal reaches no contact at all: the family that is optimal in one geometry is not merely suboptimal in another, it is disconnected.

C.3 Level three: the mechanism, which only the corner exposed

The corner contact is defined here as the boundary nodes lying within 5 % of each of the two edges meeting at the corner, a reach of 3.00 mm on the 60 mm cell. Fixing a FRACTION rather

than a length keeps the geometry invariant under mesh refinement, and fixing REACH rather than contact area is what stops the definition drifting: Appendix C.4 records two runs that reported 10.7489 and 10.6422 by extending a fixed-area contact to 53 and 52 mm, which is the full-edge geometry relabelled. Under this definition the tuned finger family reaches 9.3135 % at 22.58 % metal, which is 82.7 % of the ceiling at that metal fraction.

The corner is the hardest of the three and the only one whose loss budget inverts. Decomposing by making one conductor near-perfect at a time, shading held identical:

contact	emitter	metal	shading
centre	0.8260	0.5883	2.5055
corner	0.5885	1.7461	3.4270

At the centre, metal conduction is the smaller series loss. At the corner it is three times the emitter. That inversion generated a specific, derived prediction: minimising $\sum_b I_b^2 \ell_b / w_b$ at fixed metal area gives $w \propto I$ and a loss of $(\sum_b I_b \ell_b)^2 / \sigma A$, under which a corner-rooted tree carries 85 against rail-plus-fingers' 120 and should halve the metal loss.

The tree lost on the point contact and on a 2 mm tab, at more metal in both cases, and the derivation was recorded as a quantitative prediction that failed for a reason it could not see: the expression has no term for a minimum feature size, and a tree's leaves are each forced to 900 µm.

That much stands. What did not stand is the generalisation drawn from it. On the 5 % contact, with the Murray taper exponent swept rather than held at unity, the same topology WINS: 9.4026 % against 9.3135 % for the tuned spine, and 0.0976 against the best spine at matched metal fraction, so the margin is not bought with shading. Taper is the whole effect, the same tree scoring 6.7372 at an exponent of 0.5 and 9.4026 at 1.15. The earlier runs had swept every parameter of the tree family except the one that distinguishes it from a set of straight spokes.

The correction matters more than the number. Seven families were reported as losing to a tapered spine; what had been tested was one parameterisation of each, and what was reported was the family. Of the four retested here, the L-rail loses cleanly by 0.3062 and the multi-spine comb by 0.0770, and the comb provably CONTAINS the incumbent, since its one-spine member is the baseline exactly, so its extra spines are pure shading. The quarter fan is a tie at 0.004, which is not a result in either direction. The tree wins. A family is falsified by sweeping its parameters, not by sweeping the ones that were convenient to vary.

C.4 The three geometries are one axis

Presented as three cases, the geometries invite the reading that each is a separate problem. They are not. Holding the contact AREA fixed at 64 mm^2, so that shading is constant, and varying only how far the contact reaches along one edge from the corner, gives a single monotone curve. The cell, grid, minimum feature and layout family are identical throughout; the layout is re-optimized at every point.

reach (mm)	printed η (%)	metal	regime
2.1	9.3538	22.51%	corner
6.0	9.4394	23.05%	corner
9.9	9.6038	22.20%	corner
15.0	9.7882	21.53%	partial edge
20.1	9.9537	20.21%	partial edge
30.0	10.2685	18.96%	edge
39.9	10.5363	18.72%	edge
50.1	10.7362	16.63%	edge
60.0	10.7524	16.70%	edge, as a strip
full edge, registered busbar		16.28%	10.7952

The span is 1.4414 points and it is bought entirely by the extent of the fixed-potential boundary, not by contact area, which is held constant. Metal fraction FALLS as reach grows, from 23.05 % to 16.28 %, because a contact that reaches makes the tapered spine redundant: the layout optimum collapses to the minimum spine width once the boundary itself carries the current. The 0.0428 between the 60 mm strip and the registered busbar is the strip's own shading, which the registered boundary condition does not pay.

It is a coincidence, and a confusing one, that the solution-path span reported in the abstract for a 1.5 cm cell is also 1.44. That number is the spread across parameterization, seed, resolution and continuation schedule on ONE geometry; this one is the spread across geometries at a fixed solution path. The two are independent measurements that happen to agree to three figures, and neither is derived from the other.

This is the result the corner exercise was actually producing. Seven structural families, three searchers, a multi-scale ladder and a basin-hopping campaign moved the corner number by less than 0.05 in total. Moving the contact 8 mm further along the edge moves it by 0.20. Topology cannot buy reach.

It also sets a trap that this work fell into. Given freedom to choose the contact's footprint at fixed area, the search extends it into a long thin strip: one run reported 10.7489 from a contact reaching 53.4 mm, and another 10.6422 from one reaching 52.5 mm. Both are real numbers and neither is a corner contact. They are the full-edge geometry, rediscovered by an optimizer that was never told a corner tab must stay in the corner. A constraint that is not written down is not a constraint, and "contact area" was the wrong thing to hold fixed: the quantity that defines the geometry is reach.

C.5 What the levels have in common

At every level the informative result came from a variable the published formulation does not expose. Level one asked whether more design freedom helps and found the answer depends on level two. Level two asked what the contact is worth and could not explain itself without level three's loss decomposition. Level three's own first prediction was quantitative, derived, and wrong, and it was wrong for a reason the derivation could not see.

D What wins, by geometry

The plates below run from the published case outward. The first is the geometry and the scale the source work optimised, a busbar along the whole edge of a 1.5 cm cell. The other three hold the cell at 6 cm and move the contact: the whole edge again, then a pad at the centre, then a corner. Across the last three the contact spans 5 % of the cell edge in every case, so those

galleries can be read against each other and the only variable is where the contact sits.

Every design shown is printable: it meets a three-element minimum feature, is connected to its contact, carries no floating metal and no diagonal-only bridges, and is scored in the best legal realisation it has rather than in one arbitrary repair. Three elements is 900 µm on the 6 cm cells and 225 µm on the 1.5 cm one. Efficiencies are printed to one decimal because that is what the model resolves. The contact is drawn in red.

Table 6: One family set, four galleries. The winner changes with the geometry, and the gap between the best printable design and the best a searcher reaches from a uniform start does not close on any of them. The classical finger grid wins both edge geometries and takes no place at a corner.

cell	contact	best printed η (%)	best from a uniform start	gap
1.5 cm	whole edge	13.1686	12.4868	0.6818
6 cm	whole edge	10.9458	9.7194	1.2264
6 cm	centre pad	10.5792	9.9069	0.6723
6 cm	corner	9.4579	8.7550	0.7029

Table 7: Every seeded searcher run on the corner contact, against the best derived design (9.4026). A run is admitted to Figure 5 only if $\Delta > 0$ and that sign survives 10 % changes in β, photocurrent and metal conductance; *noise* marks a run whose improvement does not survive. Twelve of nineteen runs finished below the design they were given.

searcher	configuration	printed η (%)	Δ	verdict
MMA	tC	9.4579	+0.0553	helps
CNN	cnnfix6C	9.4512	+0.0486	helps
MMA	c5	9.4408	+0.0383	helps
CNN	c5	9.4238	+0.0213	helps
DE	c5	9.4033	+0.0007	noise
MMA	tA	9.4028	+0.0002	noise
attention	c5	9.3977	−0.0049	worse
random	c5	9.3961	−0.0065	worse
CNN	clean	9.3211	−0.0815	worse
MMA	clean	9.3184	−0.0841	worse
DE	clean	9.3167	−0.0859	worse
MMA	b64	9.3163	−0.0863	worse
random	clean	9.3143	−0.0882	worse

Figure 2 is the published cell, Figures 3, 4 and 5 the three 6 cm contacts. Table 6 summarises all four, and Tables 7, 8 and 9 list every seeded searcher run behind them.

D.1 How a design earns a panel

Two rules decide what appears. A searcher run is shown only if it beats the best conventional design on that contact and keeps that sign under 10 % changes in the diode parameter, the photocurrent and the metal conductance. On the corner contact five of nineteen seeded runs pass, two improve by an amount that does not survive the perturbation, and twelve finish below the design they started from.

Table 8: Searcher runs on the 1.5 cm published geometry, against the best derived design (12.6492). Every seeded run improves on it, which is the reverse of the 6 cm contacts, where searchers mostly fail to beat the structure they were handed. The two from-scratch entries are the control, and both now use the source's own starting density of 0.51 with the SIMP exponent held at 3: the network reaches 12.4868 on 6.51 % metal and gradient descent 12.3035 on 12.24 %. Thirty-four campaigns were excluded because their own recorded efficiency does not reproduce under this harness; they were run at 6 cm and share the contact name and grid size.

searcher	configuration	printed η (%)	Δ	verdict
MMA	repro15	13.1686	+0.5194	helps
CNN	cnnfixA	13.1645	+0.5152	helps
CNN	cnnfixB	13.1645	+0.5152	helps
CNN	repro15	13.1645	+0.5152	helps
MMA	from uniform 0.51	12.3035	n/a	from scratch
CNN	from uniform 0.51	12.4868	n/a	from scratch

Table 9: Every seeded searcher run on the full-edge busbar, against the best derived design (10.8410). One run improves on it. The other four land on 10.7952, the classical finger grid to four decimals, which is a reproduction of the published design and not an improvement on it. A further four campaigns were excluded before this table because their own recorded efficiency does not reproduce under this harness: they were run at 1.5 cm and share the contact name and grid size with the 6 cm campaigns.

searcher	configuration	printed η (%)	Δ	verdict
MMA	tD	10.9458	+0.1048	helps
MMA	b64	10.7952	−0.0458	worse
MMA	clean	10.7952	−0.0458	worse
DE	clean	10.7952	−0.0458	worse
CNN	cnnfix6B	10.7952	−0.0458	worse

Panels must then be distinct. Similarity is measured after a blur, because intersection over union is the wrong test for repeated structure: two finger grids whose pitch differs by one element share almost no metal pixel for pixel while looking identical, and three copies of one grid reached an earlier version of the corner plate that way. A design is suppressed when it looks like one already shown, whatever it scores, and ranked order keeps the better of the pair. This is why the plates carry nine, eight, seven and eleven panels rather than twelve. The count is a result: a gallery of twelve implies twelve answers, and the honest number of distinct printable designs this family set yields on a centre pad is seven.

E Model constants

symbol	meaning	value
σ_{metal}	metal sheet conductance	100
σ_{min}	emitter sheet conductance	0.02
j_L	photocurrent density	$310\,\mathrm{A\,m^{-2}}$
j_0	dark current density	-0.006
β	diode parameter	16.4
V_{bus}	busbar voltage, fixed	0.50 V
p_c	SIMP exponent on conduction	3.0
p_s	SIMP exponent on shading	3.0
R	cone filter radius	1.5 elements
	grid	200×200
	cell size	1.5 cm and 6 cm
	minimum feature	3 elements
	irradiance	$1000\,\mathrm{W\,m^{-2}}$

Element area follows from the cell size and grid, so three elements is 225 µm at 1.5 cm and 900 µm at 6 cm. The constants match the reference implementation, which is what allows the reproduction in Section 4.

F Adjoint derivation

The gradient $\mathrm{d}P/\mathrm{d}\rho$ is assembled by the adjoint method and chained back through the projection and the filter in reverse order. The physics returns $\mathrm{d}P/\mathrm{d}\rho_{\mathrm{filt}}$; the projection contributes its derivative at the current sharpness; the filter contributes its transpose, which for a cone filter is the filter itself. The result is verified against central finite differences on random designs.

G Searcher details

Gradient descent is the method of moving asymptotes (Svanberg, 1987), driven as an ask/tell stepper with a trust region on the design field. The trust radius starts at 0.03, shrinks on a rejected step and grows on an accepted one, with a floor to stop it stalling.

The learned generator follows Bhattacharya et al. (2021): a trainable latent of size 128, a dense layer to a coarse grid, then five blocks of leaky ReLU, bilinear upsampling, global normalisation, a 5×5 convolution and a per-pixel additive offset, with filters (128, 64, 32, 16, 1). Weights are Glorot-initialised. We train with Adam rather than L-BFGS, because under ask/tell a line search consumes a variable number of evaluations per step and the comparison here is stated in evaluations. One evaluation is one step for both arms.

Table 10: Centre contact, 6 cm cell, 200 × 200, minimum feature three elements. Every method is given the same formulation, the same budget per stage and, where it takes one, the same seed.

method	printed η (%)	what it varies
MMA, adaptive trust region	10.5030	40 000 pixels, adjoint
CMA-ES over the generator latent	10.5026	128 latent, no gradient
self-attention generator	10.4952	network weights, global attention
CNN reparameterization	10.4628	network weights, local convolution
annealing over rectangles	10.4415	structural moves, accepts worse
MMA judged on the printed number	10.1405	acceptance rule only
greedy search over rectangles	10.3845	structural moves, best improving

The variations swept over this baseline are listed in Section 6.

H Agent prompts and recorded reasoning

Each hypothesis carries the claim, the reason it was proposed, the screen that would decide it cheaply, and the quantity that screen reports. Proposals are recorded before the screen runs, so a refuted hypothesis stays in the record with its refutation rather than being removed. The register released with this report contains the full set.

I What an exhaustive search of one geometry looks like

A separate study reports a basin at 10.50 % on a point contact. This appendix records what was spent establishing it, because a claim that a value is hard to beat is only as good as the account of what was tried.

I.1 Eight search methods

Table 10 compares the methods on this contact.

Two of these were built to test a specific suspicion rather than to win. The attention generator answers whether the convolutional generator is starved of long-range structure: its receptive field is 52 pixels on a 200-pixel design, so a single output pixel sees a quarter of the cell through the convolutions. Given every token attending to every other, and distance from the fixed-value boundary as positional encoding, it gains 0.0015 over its warm start and does not improve again in 240 evaluations. The architecture is not the constraint.

The latent search answers whether the difficulty is that the printed objective has no adjoint. Searching the generator's 128-dimensional latent with CMA-ES against the printed number needs no gradient at all, and reaches the same basin.

I.2 Twelve design families

Table 11 lists every family swept and its best member.

Three of these families were proposed from the loss decomposition rather than by analogy, and all three were rejected by measurement rather than by argument. A contact pad tests whether spreading current in two dimensions where its density is highest beats spreading it in one: the optimum has no pad. Tapering the collectors tests the same trade one level down: the optimum has no taper. The non-uniform layout tests a consequence of the rail's accounting, that current from a distant collector pays the whole rail resistance while current from a near one pays almost none, so distant collectors should be fewer or shorter: swept over 6912 designs

Table 11: Parameterized families, each swept to convergence. The count is designs evaluated, not proposed.

family	printed η (%)	designs
power-law tapered rail	10.4973	440
non-uniform collector layout	10.4969	6912
linear tapered rail	10.4936	720
tapered rail with interleaving	10.4936	2304
contact pad and tapered arms	10.4280	864
uniform thick rail	10.4275	1200
ladder	10.2183	390
H-tree, Murray widths	7.9586	540
radial spokes	5.8194	14

the optimum is uniform.

I.3 What the agreement means, and what it does not

Roughly 25 000 designs, and everything that is not a losing family agrees to within 0.008. That is a wide basin, not a proof: the bound of that study is 12.2563 % and the basin sits at 85.7 % of it.

The unexploited term is one neck. At the hub the rail carries the cell's entire current through 2.4 mm, dissipating 7.3 % of output, and that width is optimal: measured directly, widening loses monotonically, because the shading cost of 0.420 exceeds the resistance saving of 0.289. Nothing that leaves the contact geometry unchanged appears to reach the remaining points.

J Reproduction

The scoring, selection and plotting code is released, together with the design arrays behind every figure. Each plate regenerates from its saved designs, and any design in the report can be re-scored through the same printability check used to produce it: binarise, remove metal not connected to the contact, enforce the three-element minimum by the best legal repair, then solve.

Two limits are worth stating. The generator for the parameterised family candidates is not included at this revision, so those designs are distributed as arrays rather than as code and the candidate pool cannot be rebuilt from source. And the searcher campaigns behind Section 6 are large, so the records are available on request rather than in the release.

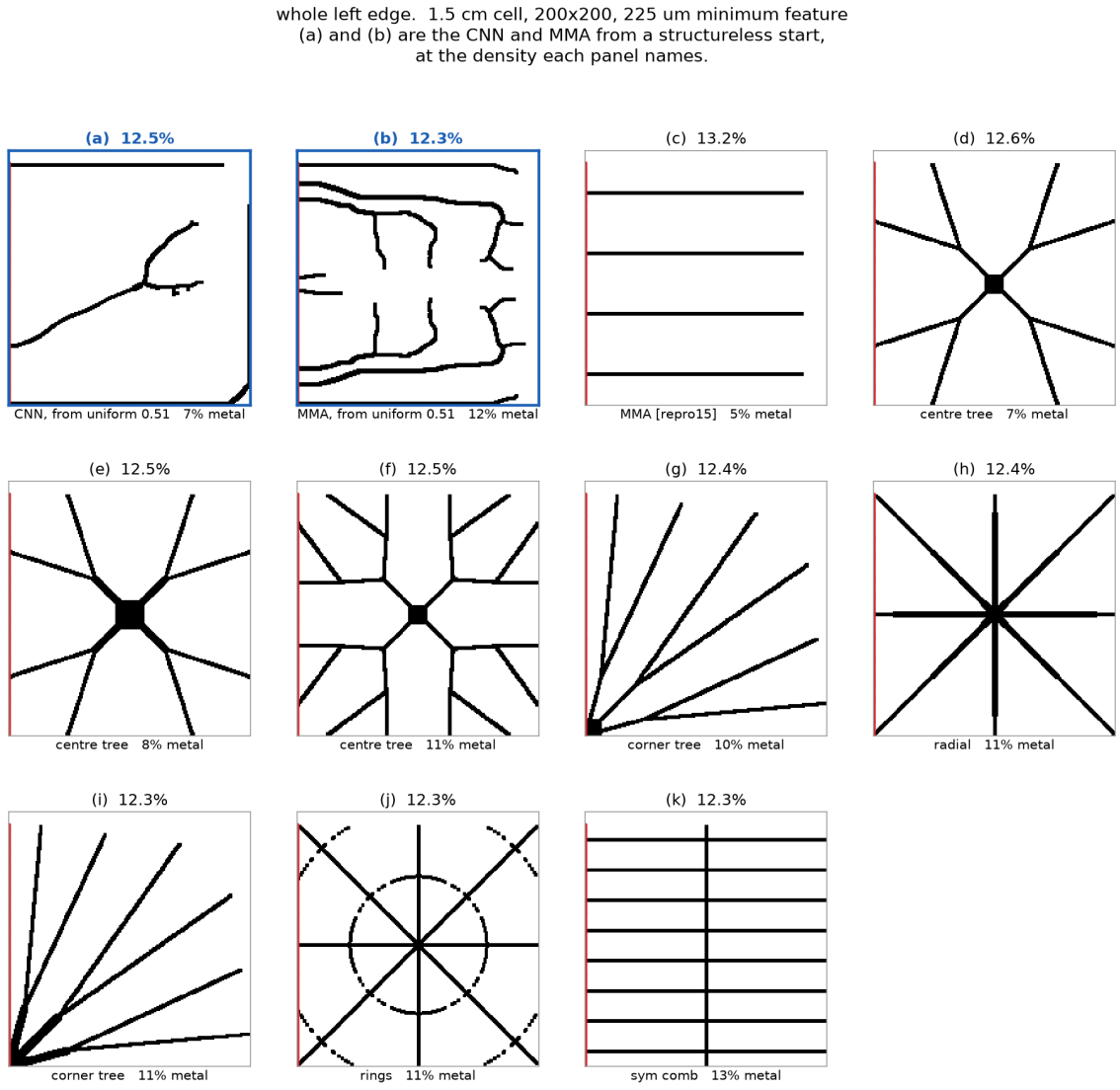

Figure 1: The published regime: a 1.5 cm cell with the busbar along a whole edge, at the grid and filter radius of the reference implementation. Panels (a) and (b) are the learned generator and gradient descent from a structureless start; (c) is the conventional finger grid, which wins on 5.43 % metal. Every design shown is binary, meets the three-element minimum feature, and is connected to the contact with no isolated metal. Panels are distinct by construction: a design is suppressed when it looks like one already shown, which is why the plate carries eleven panels rather than twelve.

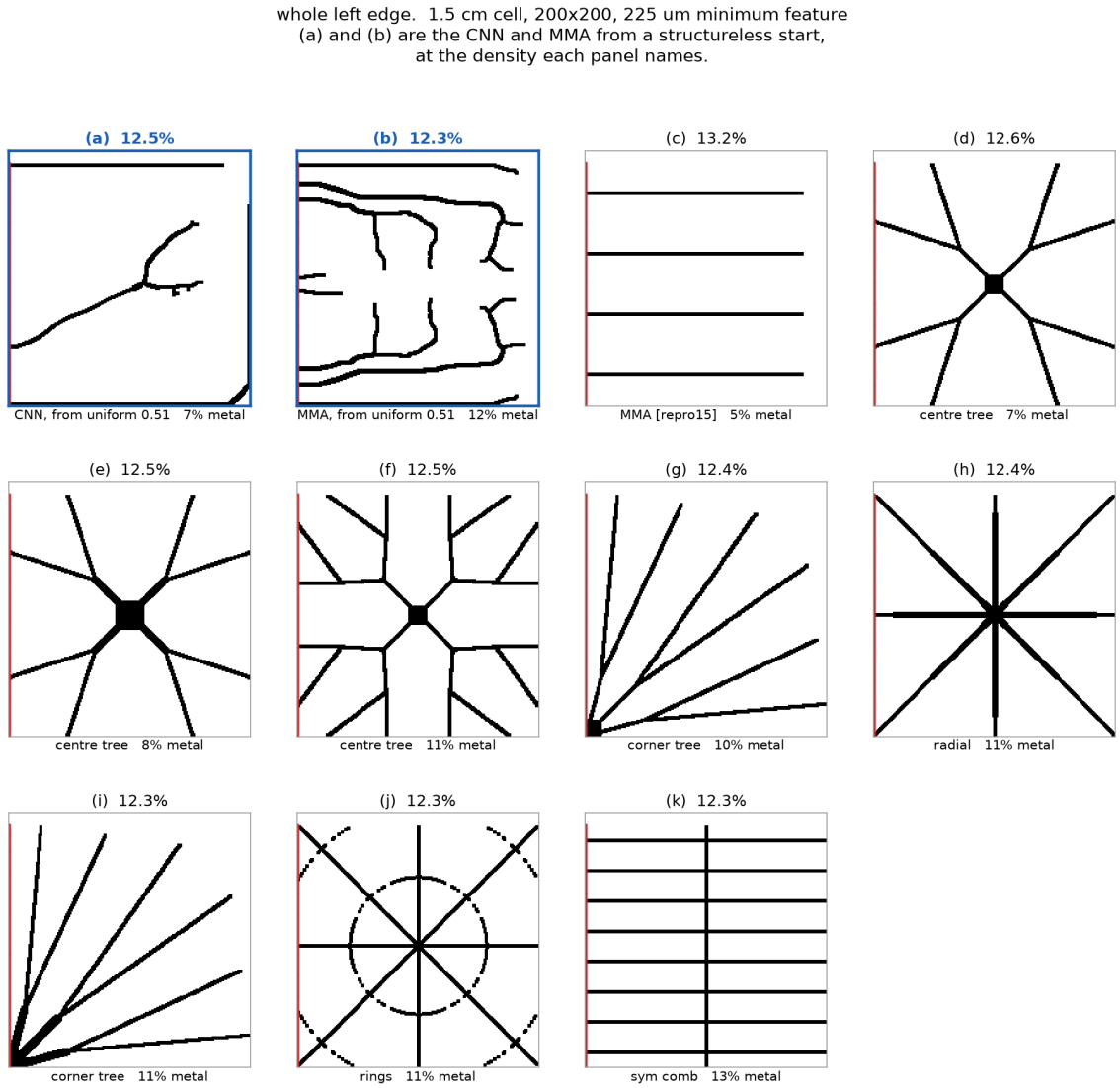

Figure 2: The published regime: a 1.5 cm cell with the busbar along a whole edge, the geometry, grid and filter radius of the source work, whose parameters we reproduce exactly. Panels (a) and (b) are the two standard methods from a structureless start at the density the source actually uses. That density is 0.51, not the 0.05 its code appears to name: the variable it initialises is a pre-sigmoid logit, so the optimisation begins at half metal and carves material away. With that start, with the SIMP exponent held at 3 from the first iteration rather than ramped, and with no Heaviside projection at all, the design binarises on its own to a greyness of 0.008: penalising conduction as ρ^3 and photocurrent as $(1-\rho)^3$ makes an intermediate density useless as a conductor and nearly opaque at once, so the optimiser vacates it without being told to. The network reaches 12.4868 on 6.51 % metal and gradient descent on the field reaches 12.3035 on 12.24 %, both fully compliant with the three-element rule. Neither reaches panel (c), the classical finger grid, at 13.1686 on 5.43 %. All twelve designs here satisfy the minimum feature exactly: every metal element lies inside a 3×3 block of metal, checked by an opening that is zero-padded, so a rail lying on the cell boundary is eroded from outside rather than credited with material it does not have.

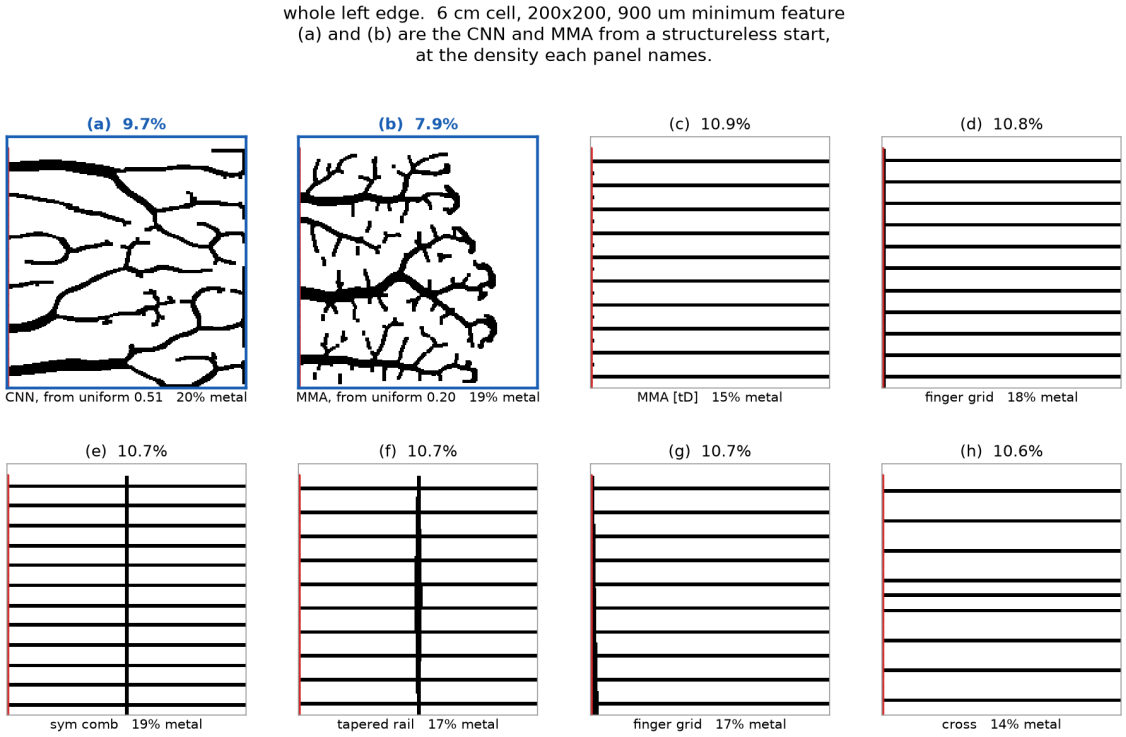

Figure 3: Full-edge busbar, under the same admission and distinctness rules. The published topology is confirmed and its published proportions are not. Panel three, the best design on this geometry, is a set of parallel fingers carrying no spine, exactly the form the classical argument predicts: with the busbar along a whole edge every point already has a direct path out through the emitter, so routing is free and the only remaining task is to minimize mean distance to metal at fixed metal area, which parallel lines solve exactly. But it reaches 10.9458 on 15 % metal, where the classical grid of panel five reaches 10.7914 on 18 %: more efficiency on a fifth less silver, from the same family at a different pitch. A symmetric comb, panel four, also beats the plain grid. Four separate searcher runs converged to 10.7952, the classical grid to four decimals, which is a reproduction of the published result rather than an improvement on it, and eight of nine seeded runs finished below the best derived design. So the literature's claim about this geometry is right about the shape and wrong about the numbers, which is the kind of error a family sweep finds and a single reported optimum hides.

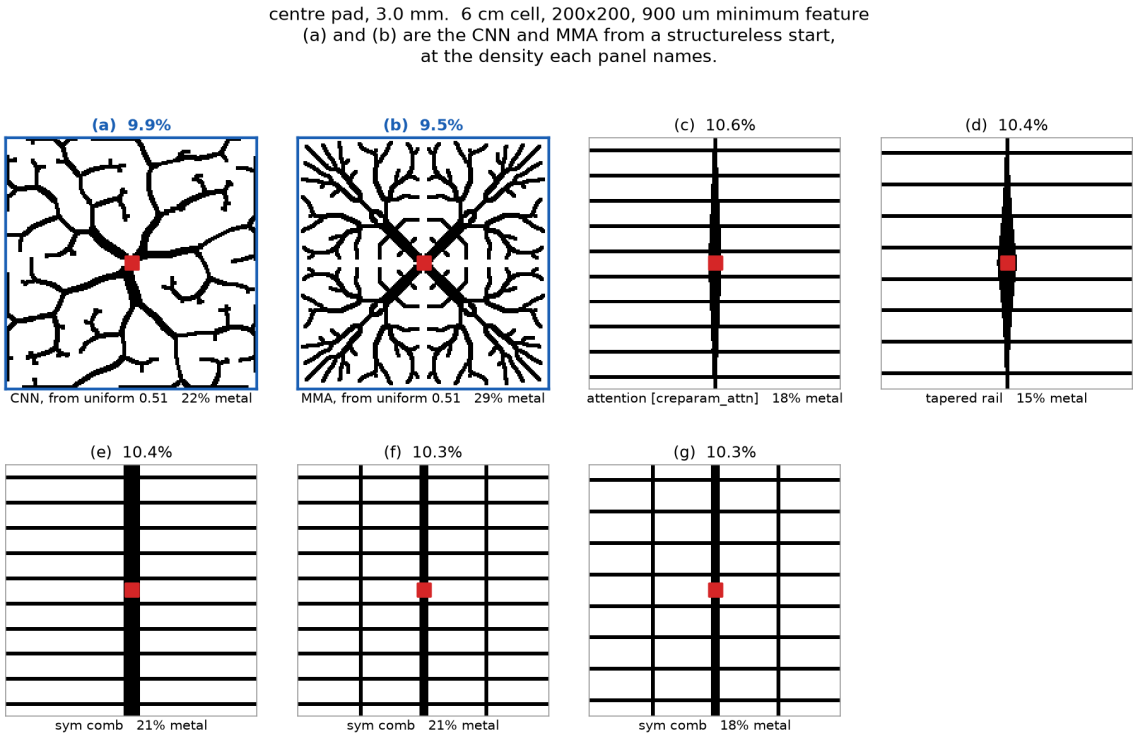

Figure 4: Centre contact, the same 5 % rule applied at the middle of the cell, and the clearest negative result in the set. One searcher design earns a panel, the attention reparameterization at 10.5792, which beats the best derived design by 0.0015 on a quantity a 1 % change in β moves by 0.18. Two further runs improved on the incumbent and were then suppressed as near-duplicates of the tapered rail they had refined, so the honest summary of this contact is that no searcher produced a design that is both better and different. Panels four onward are tapered collectors and symmetric combs: the rail is widest at the pad and narrows to the minimum feature at the cell edges, because the current it carries falls linearly. A contact with a symmetry needs a family set that can express that symmetry. Without one, every design on this plate rooted its collector at the left edge and gathered current that then had to cross the whole cell. Adding a symmetric family was worth 0.6973; every searcher run since is worth 0.0015 on top of it.

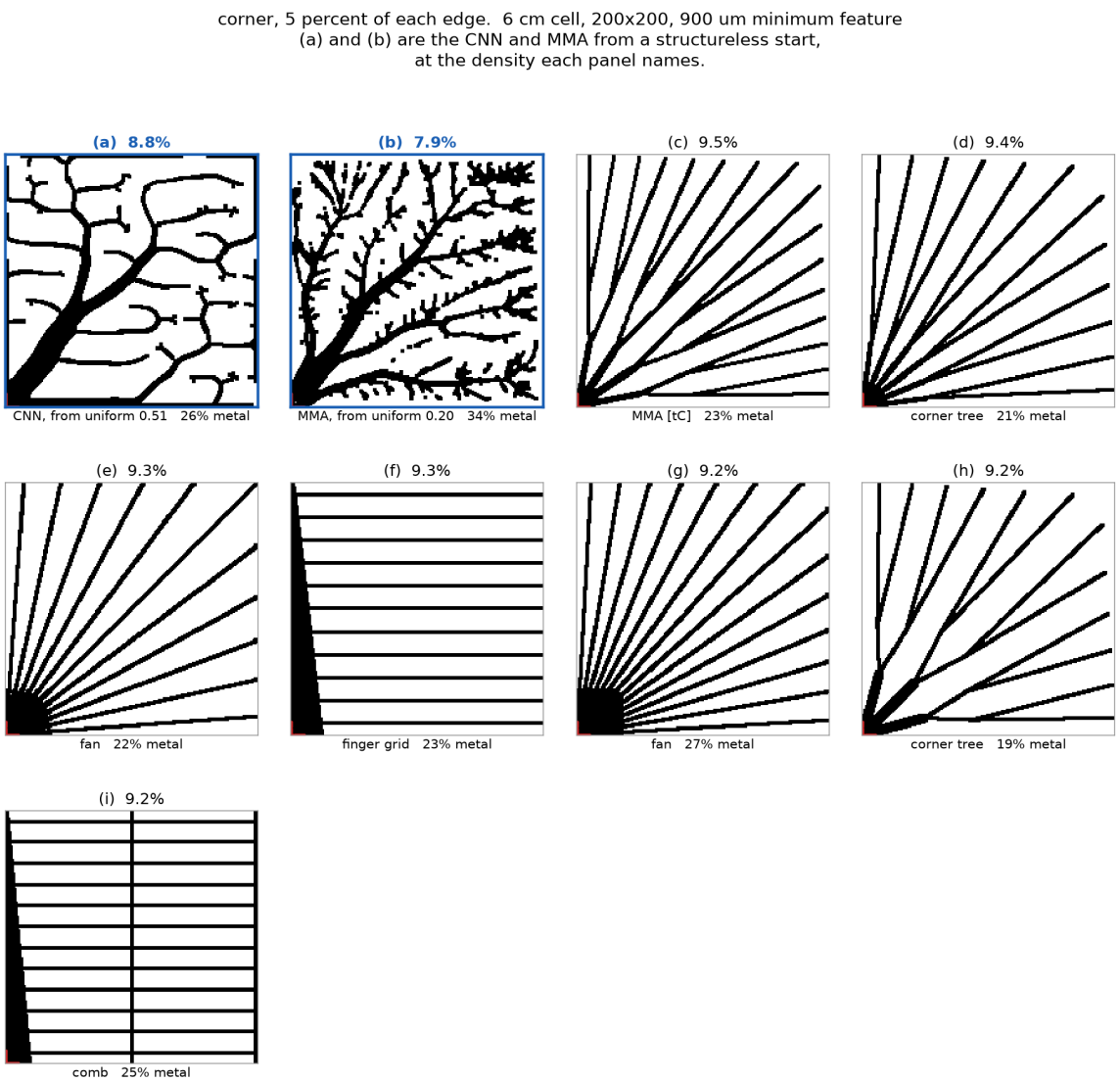

Figure 5: Corner contact. Efficiencies are shown to one decimal because that is what the model resolves: a 1 % change in the diode parameter β moves the absolute efficiency by 0.18, which is three times the largest improvement any searcher achieved. Panels one and two are reserved rather than ranked, being the two standard methods started from a uniform density of 0.20 with no structural prior; they finish last, and spend 32 % and 34 % of the cell on metal doing it against 23 % for the design that wins, so the deficit is not a metal budget they were denied. MMA's is the more interesting failure: from a structureless field it recovers a branching dendrite, the correct morphology for a corner contact, and then fails on length scale rather than on topology. A searcher run is admitted only if it beat the best derived design and kept that sign under 10 % changes in β, photocurrent and metal conductance; the rest are in Table 7 rather than deleted. Panels are then required to be distinct, and the plates are shorter than twelve because of it. Similarity is measured after a blur: intersection over union is the wrong test for periodic structure, since two finger grids whose pitch differs by one element share almost no metal pixel for pixel while looking identical, and three copies of one grid reached an earlier version of this plate that way. Blurring first collapses the pitch difference and compares the arrangement, which is what a reader actually sees. A design is suppressed when it looks like one already shown, whatever it scores: two designs a reader cannot tell apart do not become distinct because their numbers differ, and ranked order means the better of a pair is the one kept. What the freed slots reveal is the point of the plate: corner-rooted trees at several spoke counts, a fan on a solid hub, finger grids and combs, spanning 0.28 in efficiency and 8 % in metal. The tree is the result and the searchers are polish on it, which is the reverse of the emphasis the method literature would suggest. An earlier pass of this work reported the ranking inverted, having swept every parameter of the tree family except the Murray taper exponent that distinguishes it from a set of straight spokes.