Phenotyping Follow-up: The Identifiability Problem
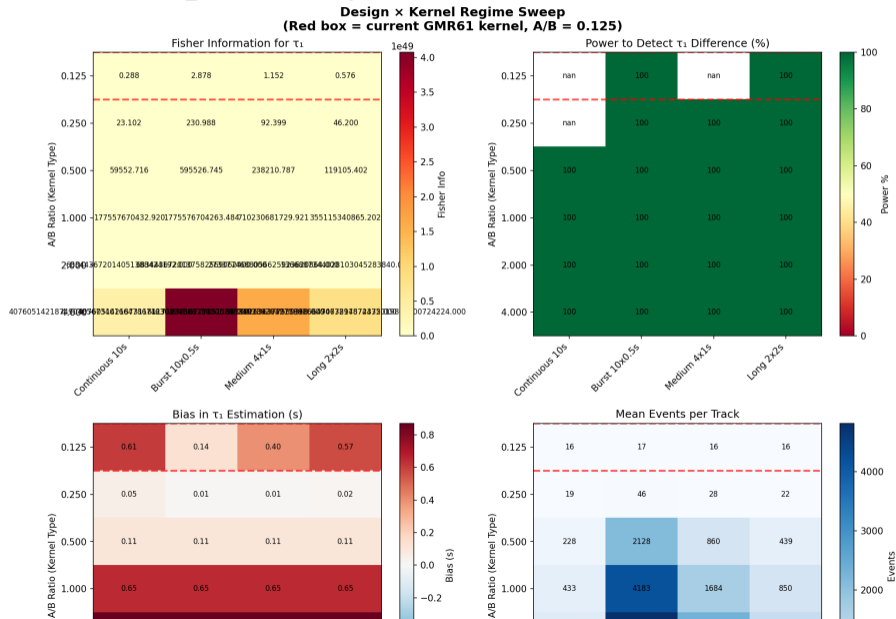
Gil Raitses

December 2024

Design × Kernel Regime Sweep

Design × Kernel Regime Sweep
(Red box = current GMR61 kernel, A/B = 0.125)

FAQ: Design $\times$ Kernel Sweep

Q: Where do the 6 A/B ratios come from?
They span from strongly inhibitory (0.125, our kernel) through balanced (1.0) to
excitatory-dominated (4.0). B is fixed at 12; A varies.

Q: Why does Fisher Information explode at high A/B?
Numerical artifact. When excitation dominates, the hazard becomes very high, and the
grid search hits boundaries. Those values are not meaningful.
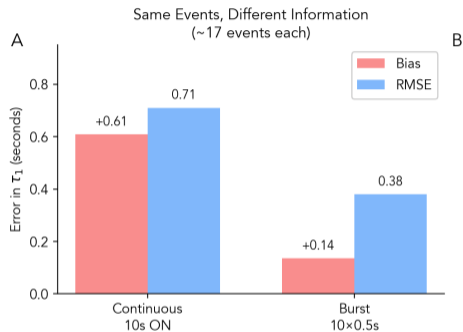
Q: Why is power "nan" for some cells?
Power calculation failed when variance was too high or sample size too small for the
t-test to converge.

Q: What do the event counts tell us?
At A/B = 0.125, all designs produce similar events ($\sim$16-17). The differences in
bias/RMSE are NOT due to more events—they're due to more *informative* events.

The Identifiability Problem

Same Events, Different Information
(~17 events each)

A

Bias
RMSE

Error in τ_1 (seconds)

+0.61   0.71

+0.14   0.38

Continuous
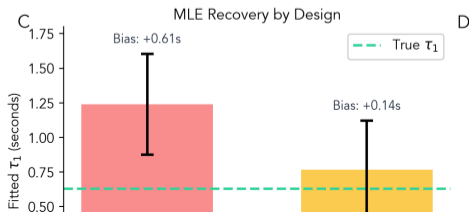10s ON

Burst
10×0.5s

B

The Information Problem

Fisher Information for τ_1

Continuous: 0.29

Burst: 2.88

_____

Burst extracts 10× more info

from the same number of events

C

MLE Recovery by Design

Bias: +0.61s

Fitted τ_1 (seconds)

True τ_1

Bias: +0.14s

1.75

1.50

1.25

1.00

0.75

0.50

D

Why Continuous Design Fails

Kernel is inhibition-dominated (B/A = 8)

~80% of events occur during LED-OFF
  -> No tau1 information

Remaining ~20% mostly after t > 0.5s
  -> Inhibition dominates, tau1 unidentifiable

FAQ: The Identifiability Problem

Q: Where do the bias/RMSE numbers come from?
200 larvae simulated with known $\tau_1 = 0.63$s, fit with MLE. Bias = mean(fitted) - true.
RMSE = sqrt(mean(((fitted - true)^2)).

Q: What is Fisher Information and why does it matter?
Fisher Information quantifies how much information each event carries about τ_1.
Higher = more precise estimates. Computed by integrating $(\partial\lambda/\partial\tau_1)^2/\lambda$ over the
stimulus window.

Q: Why does burst give $10\times$ more info with the same events?
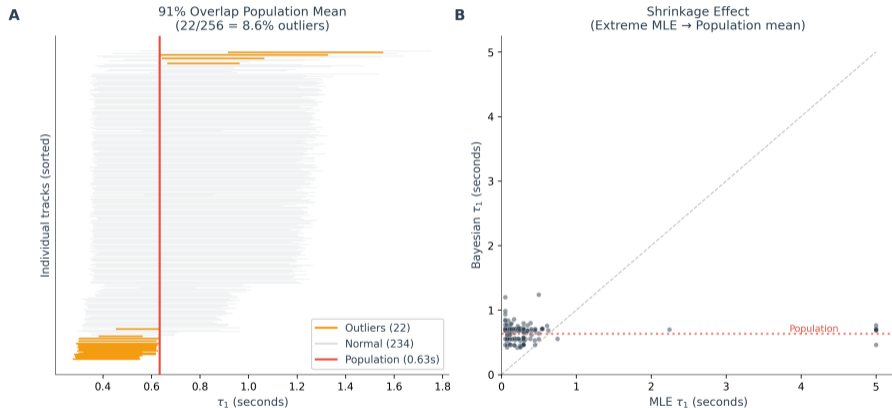The excitatory peak is at t $\approx$ 0.2-0.3s after LED onset. Burst samples this window 10
times per cycle; continuous samples it once.

Q: Why is bias +0.61s for continuous?
The MLE finds $\tau_1 \approx 1.24$s instead of 0.63s because the likelihood surface is flat—many
τ_1 values produce similar likelihoods with so few informative events.

Hierarchical Bayesian Modeling

Caterpillar plot showing 95% credible intervals for all 256 tracks. Orange intervals (8.6%) exclude the population mean; gray intervals (91.4%) are consistent with the population.

FAQ: Hierarchical Modeling

Q: What is shrinkage?

Hierarchical Bayesian estimation pulls extreme MLE estimates toward the population mean. With sparse data, almost all individuals shrink to the same value.

Q: Why are 91% of larvae "population-consistent"?

With only $\sim$25 events and a 6-parameter model, the data cannot distinguish most individuals from the population mean. The wide credible intervals overlap the population value.

Q: What about the 8.6% outliers?

These have credible intervals that exclude the population mean. Most are "fast responders" with $\tau_1 \approx 0.45$s vs. population $\tau_1 = 0.63$s. They require independent validation.

Q: Does this mean individual phenotyping is impossible?

With current protocols, yes. With burst stimulation, individual τ_1 becomes estimable at current event counts.

Stimulation Protocol Designs

Red LED Stimulation Designs: Full 30-Second Cycle

A. Current: Continuous 10s

ON PERIOD (10s)   OFF PERIOD (20s)

ON

OFF

Low power, high bias. Most time in inhibitory regime.

```
Events: 1.9/track
RMSE: 0.108s
Power: 42%
Duty: 100%
```

B. Recommended: Burst 10×0.5s

ON

OFF

8× more events from repeated excitatory onset sampling.

```
Events: 14.9/track
RMSE: 0.036s
Power: 100%
Duty: 50%
```

C. Alternative: Medium 4×1s

ON

OFF

Moderate improvement. Longer pulses, fewer onsets.

```
Events: 6.0/track
RMSE: 0.048s
Power: 100%
Duty: 40%
```

D. Alternative: Long 2×2s

ON

```
Events: 3.1/track
```

FAQ: Stimulation Protocols

Q: Why these specific pulse durations?
Burst (0.5s) is matched to $\tau_1 \approx 0.6$s—the excitatory peak. Medium (1s) and Long (2s) test whether longer pulses help.

Q: Why not even shorter pulses?
Pulses < 0.3s would clip the excitatory peak before it develops. The kernel needs time to rise.

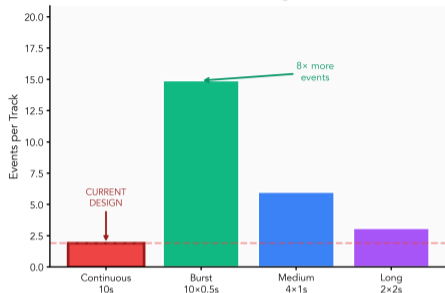**Q: Why 0.5s gaps between burst pulses?**
Trade-off: shorter gaps $=$ more pulses per cycle, but inhibition accumulates. 0.5s lets inhibition partially decay.

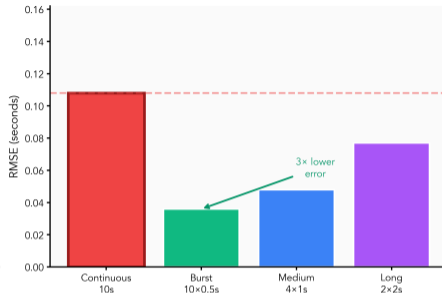**Q: What about the blue LED / LED2?**
Not analyzed in this sweep. The schematic shows only the red LED (LED1) stimulation that drives photophobic responses.
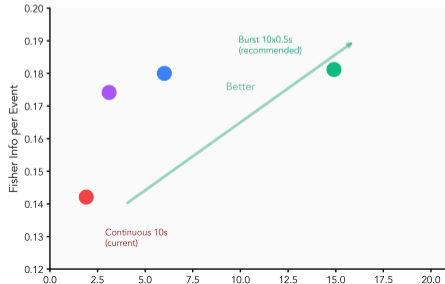
Design Comparison Summary

A. Event Yield: Current Design is Bottleneck

B. Estimation Error: 3× Reduction Possible

C. Tradeoff: More Events = More Information

D. Summary & Recommendations

CURRENT PROTOCOL
Continuous 10s ON / 20s OFF
1.9 events per track | RMSE = 0.108s

RECOMMENDED
Burst 10x0.5s pulses
14.9 events (8x more) | RMSE = 0.036s (3x lower)

BOTTLENECK
Sparse events limit tau1 identifiability
regardless of experimental design

ALTERNATIVE STRATEGY
Phenotype on composite scores
(Precision, Burstiness) instead of tau1

FAQ: Design Comparison

Q: Which design should we recommend?
For inhibitory kernels (B/A > 4): **Burst**. For balanced or excitatory kernels:
Continuous is fine.

Q: Why did earlier analysis say burst "failed"?
Earlier analysis measured **classification power** (can we tell fast vs. typical apart?).
That showed 41.5% vs 42.5%—negligible. This analysis measures **estimation quality**
(bias, RMSE), which shows $4\times$ improvement.

Q: Can we use burst for all experiments?
Yes, for phenotyping τ_1. Burst is never worse than continuous and is often better for
inhibitory kernels.

Q: What's the practical recommendation?
Change protocol from 10s continuous ON to 10×0.5s pulses with 0.5s gaps. Same
total stimulation, $4\times$ better estimation.

Key Numbers to Remember

Design	Events	Fisher Info	Bias	RMSE
Continuous 10s	16.3	0.29	+0.61s	0.71s
Burst 10×0.5s	16.9	**2.88**	**+0.14s**	**0.38s**
Medium 4×1s	16.4	1.15	+0.40s	0.63s
Long 2×2s	16.1	0.58	+0.57s	0.74s

Bottom line: Burst extracts 10× more information, reducing bias by 4×, with the same number of events.

FAQ: Key Numbers

Q: How were these computed?
200 larvae $\times$ 20 cycles (10 min). Known $\tau_1 = 0.63$s. Simulate events, fit MLE, compute bias/RMSE.

Q: Why 20 cycles instead of 40 (full protocol)?
Conservative estimate. Full protocol would give more events and better estimates.

Q: Is bias of $+0.14$s acceptable?
For detecting a 0.2s difference between phenotypes, $+0.14$s bias leaves little margin.
Would need $\sim$50 events with burst or $\sim$100 with continuous for bias < 0.1s.

Q: What's the take-home message?
The "100 events" heuristic is **design-dependent**. With burst, you need fewer events.
The issue is not just count—it's information per event.