BASIC CONCEPTS · RUN EVERY IDEA IN THE PAGE

The line makes less than its machines promise.

Not a report about simulation — a simulation you can run. Turn the knobs, break the line, and watch the theory hold.

Live 4-station line · sampling real process times

0 pcs/h vs 1200 pcs/h ceiling

EDITION 2 · 2026 · RASQINTELLIGENCE · BUILT WITH NO FRAMEWORKS

01 / WHY SIMULATE

The line is not the sum of its machines

A spreadsheet adds average rates and hands you a number. A real line never reaches it. Here is why — you can watch it happen.

Give four machines average rates and a spreadsheet will tell you the line runs at the

speed of its slowest one. That number is a *ceiling*, not a forecast. The moment process times **vary** — and they always do — stations start to block (the buffer ahead is full) and starve (the buffer behind is empty). Every block and every starve is throughput you planned for and never got.

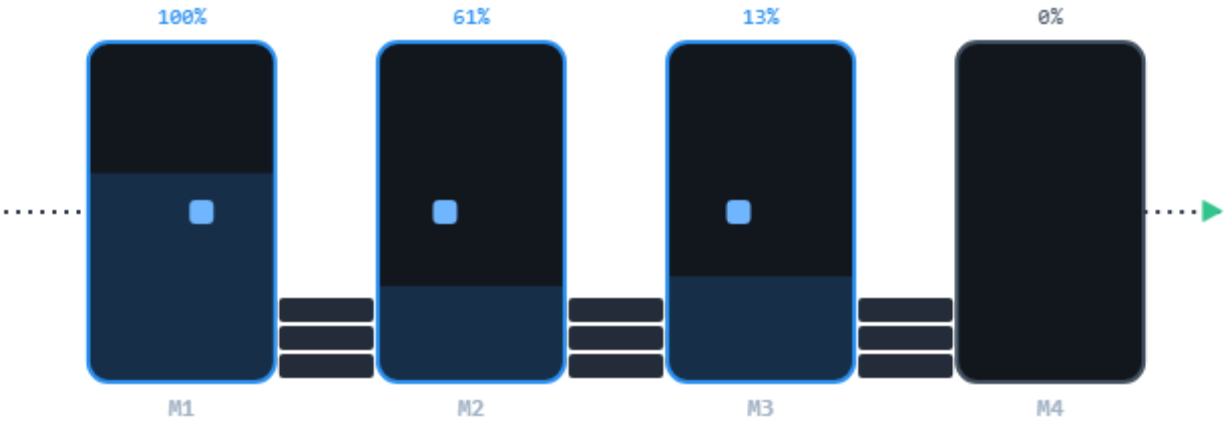
The line below is a real discrete-event simulation running in your browser. Press **Run**. Then drag **variability** up from zero and watch the gap between the dashed ceiling and the live rate open.

EXPLORABLE · DISCRETE-EVENT SIMULATION

The Living Line

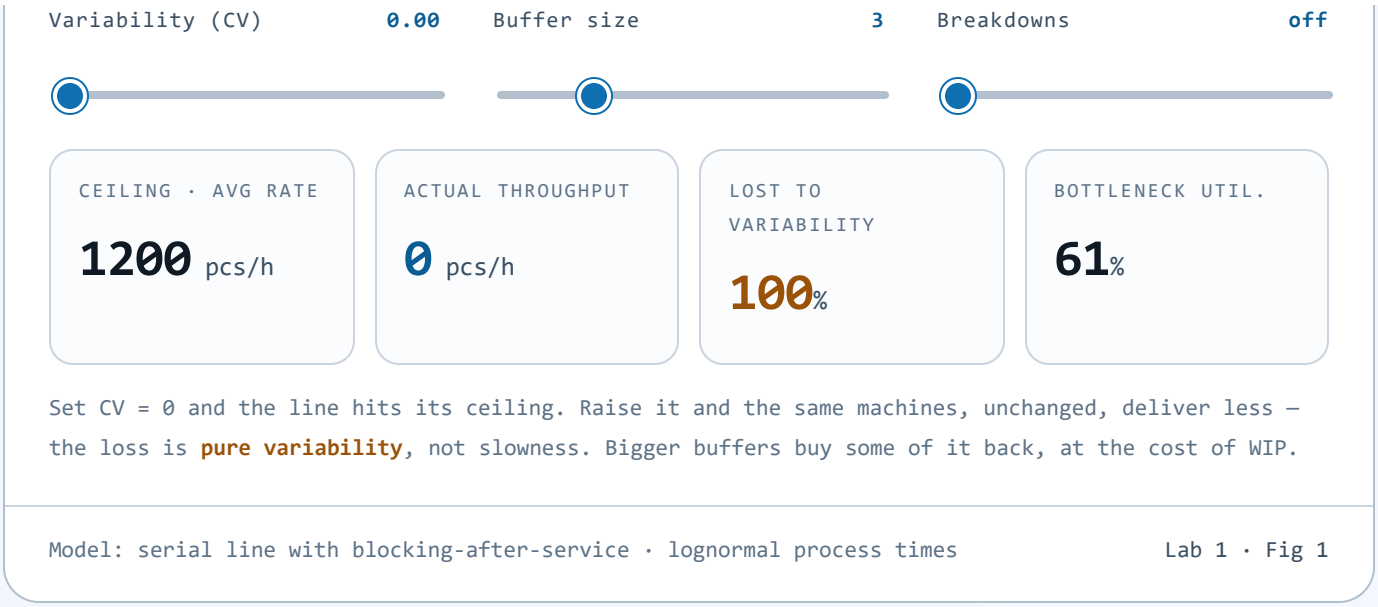
Four stations · buffers between them · process times drawn from a distribution

Live output — trailing throughput, last 50 samples vs the dashed ceiling



working · blocked (buffer full ahead) · starved (buffer empty behind) · breakdown · finished part

Pause ↺ Reset 1× 3× 8×



Discrete-event simulation advances time event by event — a part finishes, a machine breaks, a setup starts — and tracks every station, buffer and part. That is why it captures what averages miss, and why it can answer the only question that matters before you spend money: **what will this line actually make, where is it constrained, and what is a change worth in dollars per year?**

> **The purpose of simulation is not a prettier number — it is an honest one, early enough to change the decision.**

Sources: Hopp & Spearman, *Factory Physics* (3rd ed.); Law, *Simulation Modeling and Analysis*.

02 / THE CENTRAL VILLAIN

Variability, not slowness, steals throughput

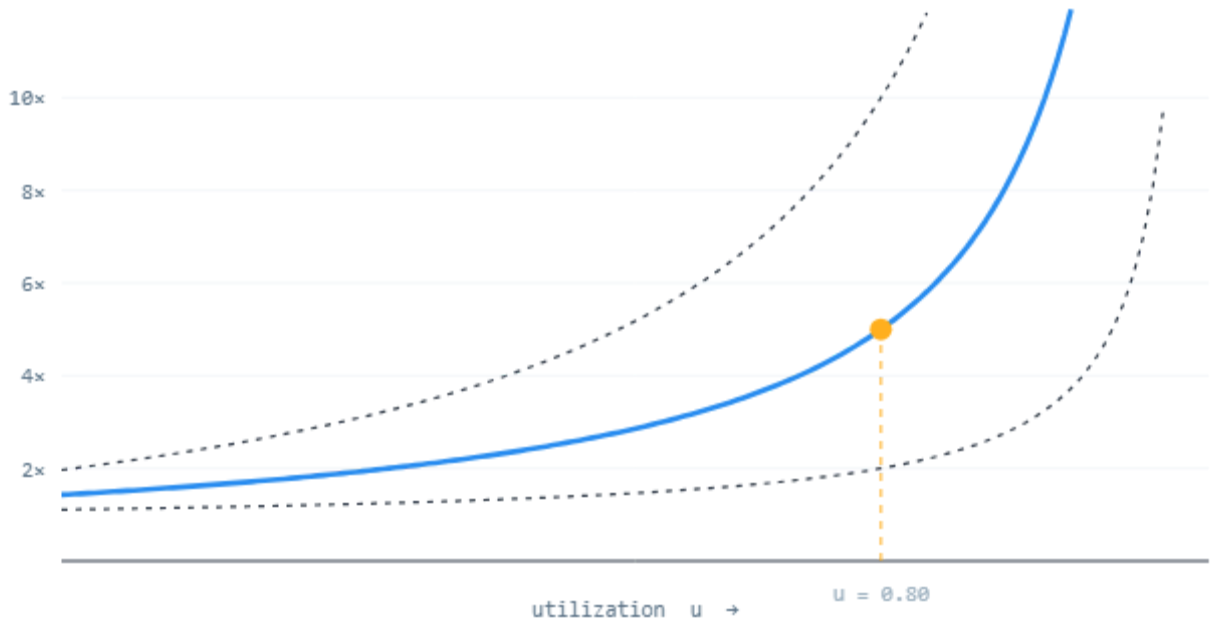
If you learn one idea from this handbook, make it this one. And you can feel it: push utilization toward 100% and cycle time doesn't rise — it explodes.

Measure variability with the coefficient of variation, **c = σ / μ** — the spread relative to the mean. Kingman's famous VUT equation says the queue time at a station is variability × utilization × time. Drag the two knobs below. Notice two things: as utilization **u → 1** the curve goes vertical (a non-bottleneck run flat-out is self-sabotage), and lifting variability lifts the *whole* curve.

EXPLORABLE · KINGMAN VUT

Where the queue explodes

Cycle time (× base) = 1 + (c²/1)·(u/(1−u)) — waiting line grows to the right of the marker



Utilization u 0.80

Variability c 1.00

CYCLE TIME	OF WHICH, QUEUE	WIP PILED UP (LITTLE)
5.0× base	4.0× base	4.0 parts

Try holding c = 1 and sliding u from 0.80 to 0.97: cycle time roughly **quadruples** for the same

machine. That is the cost of chasing utilization on a variable line.

Kingman (1961) G/G/1 approximation · the quantitative case for SMED, TPM, Six Sigma Lab 2 · Fig 2

> **Cutting variability cuts cycle time linearly — often cheaper than buying capacity.**

This is why a perfectly balanced line performs worse than an unbalanced one with a clear bottleneck: the balanced line has no slack to absorb variation. Chase the *spread*, not only the average.

Sources: Kingman (1961), G/G/1 waiting-time approximation; Hopp & Spearman, law of variability buildup.

03 / THE LAWS OF FLOW

More WIP buys lead time, not output

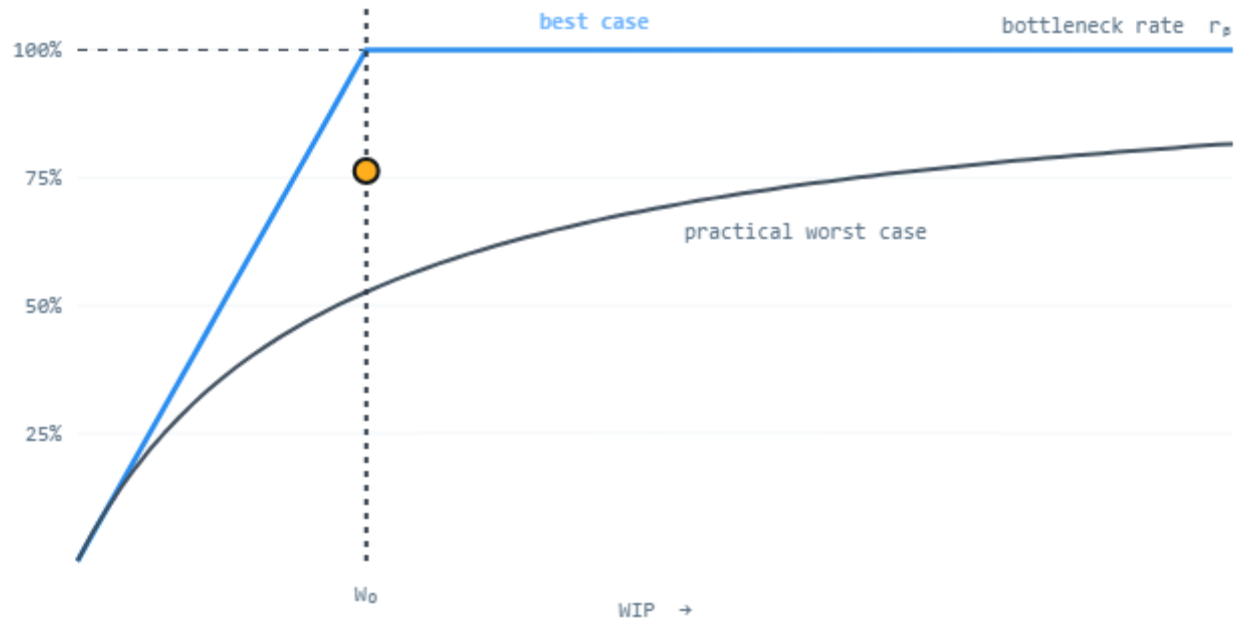
Three relationships govern every line.
They always hold, and a trustworthy
simulator must reproduce them.

Little's Law ties the three numbers you care about: **WIP = Throughput × Cycle Time**. There is a critical WIP, W_0, the exact amount a variability-free line needs to hit full speed at minimum cycle time. Slide WIP below and watch the operating point ride between the best-case hockey-stick and the practical-worst-case curve — the distance between them is what you lose to variability. Push past W_0 and throughput barely moves while lead time climbs.

EXPLORABLE · LITTLE'S LAW & THE CONWIP CURVE

Throughput vs work-in-process

Best case rises to the bottleneck rate at W_0, then flattens — real lines sit below it



WIP cap (CONWIP) **10**

THROUGHPUT	LEAD TIME	PAST W_0?
76% of r_β	**1.31**× min	**no**

Left of W_0 you are starving the line; right of it you are just **aging inventory**. The knee is where a line should run.

Little (1961); Goldratt, *The Goal*; Hopp & Spearman best-case / practical-worst-case — Lab 3 · Fig 3

04 / THE LINE'S LIFE

New, mature, aging — not one rate

A line doesn't perform the same in its

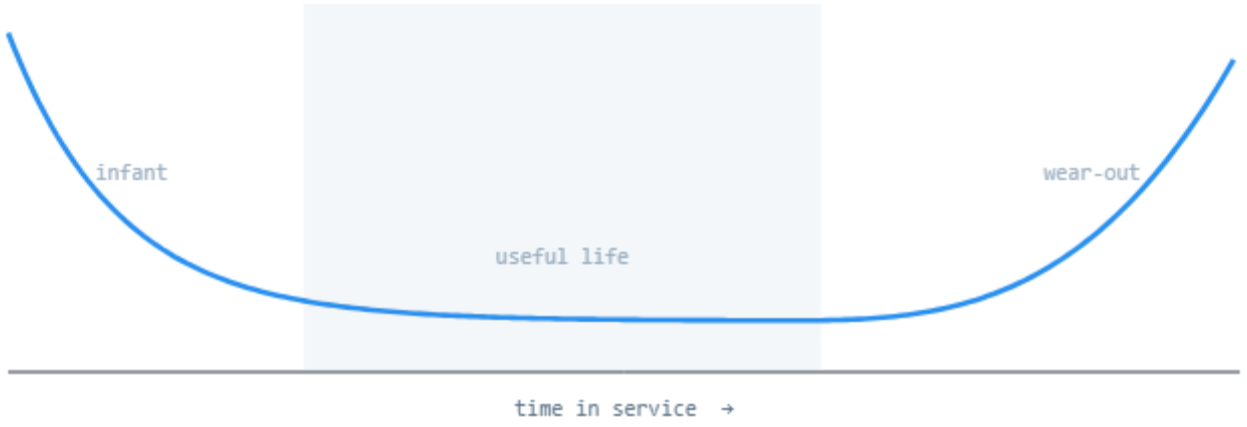
first month, its third year, and its tenth. Using one number for all three is a common — and expensive — mistake.

Reliability follows the **bathtub curve**: high early failures (infant mortality), a long flat middle (useful life), and a rising tail (wear-out). Productivity follows the **learning curve**: cycle time falls as experience grows. Match the model to the line's age and complexity — ramp-up for a launch, steady-state for a mature plant, wear for an old asset.

REFERENCE · RELIABILITY + LEARNING

Failure rate follows age, not a single rate

Bathtub hazard mapped to three lifecycle regimes



Reliability engineering; Weibull (1951); Wright (1936) learning curve

Fig 4

05 / DOWNTIME IN THREE TIERS

The stops nobody logs cost the most

"Downtime" is not one thing. The short

stops nobody records often cost more than
the breakdowns everybody remembers.

Losses live on three time scales: rare long **breakdowns**, occasional **stress-coupled** events whose rate rises when you push the machine, and the many tiny **micro-stops** — a sensor re-trip, a part re-seat — that never make it into a log yet silently erode the constraint. Watch one line's events tick past, by scale:

LIVE · EVENT TIMELINE

Same line, three time scales

Breakdowns · stress events (NHPP) · micro-stops — running in real time

breakdown stress event micro-stop

BREAKDOWN DOWNTIME	MICRO-STOP DOWNTIME	MICRO ÷ BREAKDOWN	TOTAL DOWNTIME
82.7%	9.3%	0.11×	103.7%

Each micro-stop is trivial; there are just so many of them. Watch the two shares converge — the losses **nobody logs** rival the breakdowns everybody does.

Nakajima, *Introduction to TPM* (1988), the six big losses; NHPP Fig 5

Garbage in, garbage out

A simulator is only as trustworthy as its inputs. The discipline isn't "get perfect data" — it's "know how good your data is, and say so."

Process times are usually **right-skewed** (lognormal or gamma); time-to-failure is Weibull; truly random events are exponential. Defaulting to a *normal* distribution for a skewed process invalidates everything downstream. Fit the distribution to data, sample enough across shifts and products, and label numbers as estimates until real data replaces them.

REFERENCE · DISTRIBUTION FIT

Real process times are right-skewed

Cycle-time histogram with a fitted lognormal (Kolmogorov–Smirnov)

Law, *Simulation Modeling and Analysis*; Kolmogorov–Smirnov goodness-of-fit

Fig 6

One run is an anecdote

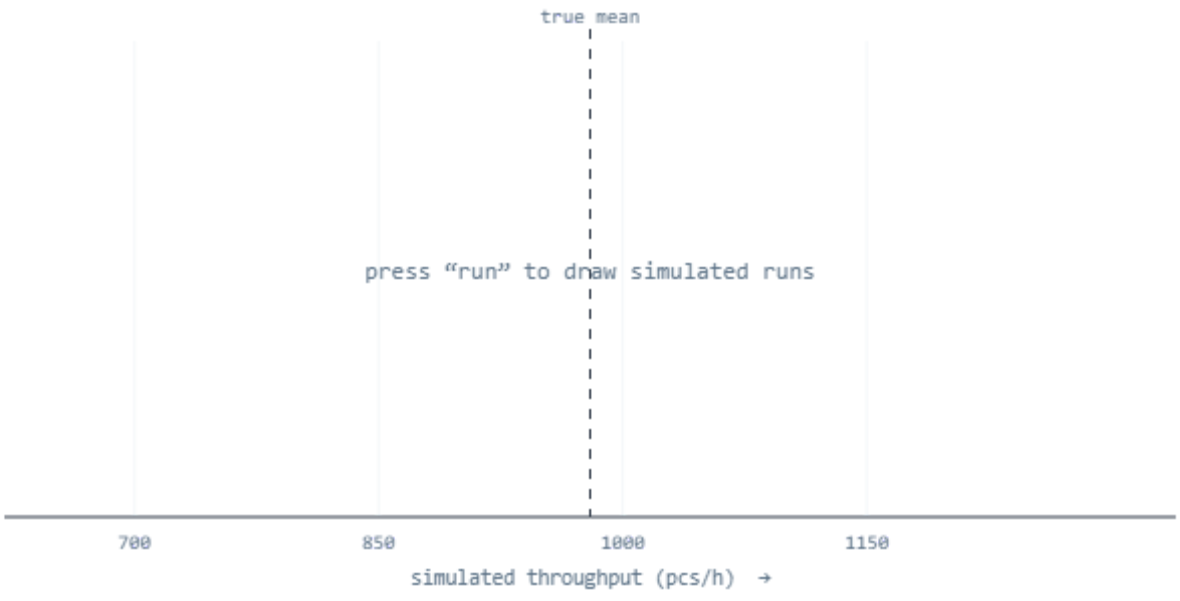
A single simulation run is a single draw from a distribution. Turning it into evidence takes replications — the same discipline that separates a study from a demo.

Press **run one replication** a few times: each is a different number. Now hold it down — as replications accumulate, the mean settles and the 95% confidence interval narrows. You choose how tight the decision needs to be. Never decide on one run; the interval is the difference between "we think" and "we measured."

EXPLORABLE · MONTE CARLO OUTPUT ANALYSIS

Replications tighten the estimate

Each dot is one simulated run; the band is the 95% confidence interval on the mean



▶ Run one replication ▶▶ Run 25 ↺ Reset

REPLICATIONS N	MEAN ESTIMATE	RUN-TO-RUN Σ	95% CI HALF-WIDTH
0	— pcs/h	— pcs/h	— pcs/h

The wide grey bell — the spread of single runs — barely changes. The blue bell is the distribution of the **mean estimate**: it narrows as **1/√n**, so quadrupling the runs halves your uncertainty.

Law, *Simulation Modeling and Analysis* — output analysis, replications, CRN — Lab 7 · Fig 7

---

08 / FINDING WHAT MATTERS

Design of experiments beats guessing

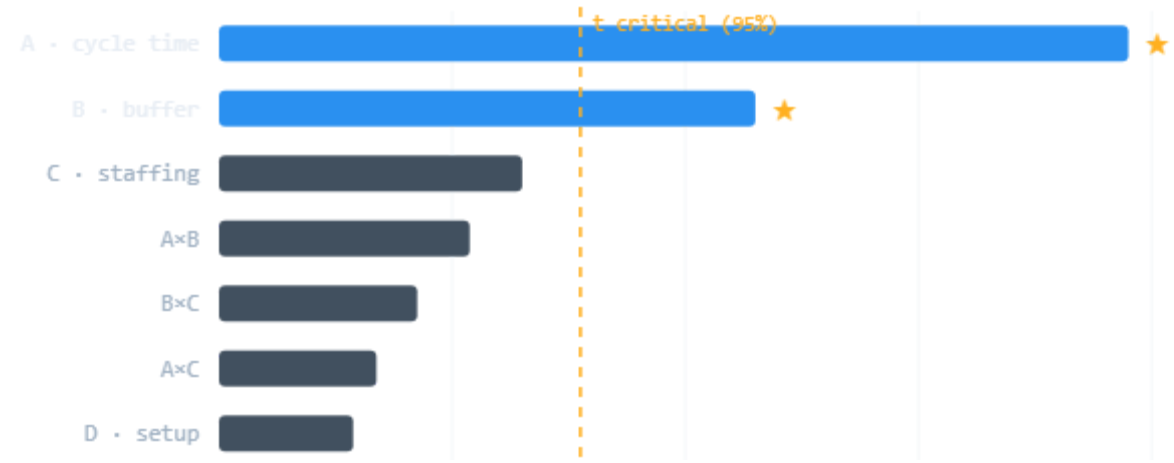
Changing one knob at a time is slow and blind to interactions. A factorial design finds which factors actually move the line — and where they stop helping.

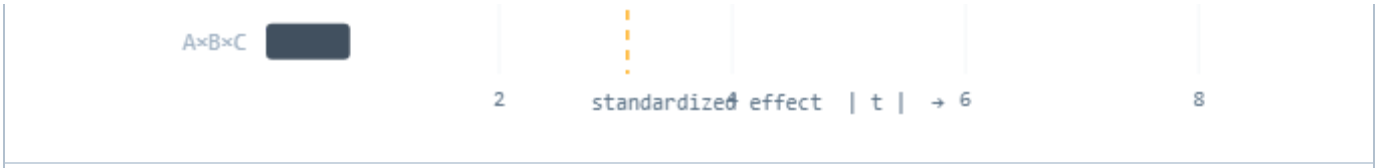
A standardized-effects **Pareto** ranks factors against a significance threshold, so you spend effort on the two levers that matter instead of the five that don't. Push a lever far enough and the return flattens — the constraint moves elsewhere.

REFERENCE · FACTORIAL DOE

Two factors move the line; the rest is noise

Standardized effect |t| per factor, ranked against the 95% threshold





Montgomery, *Design and Analysis of Experiments*; Goldratt, five focusing steps Fig 8

09 / READING THE MONEY

"The line is slow" becomes "$X per year"

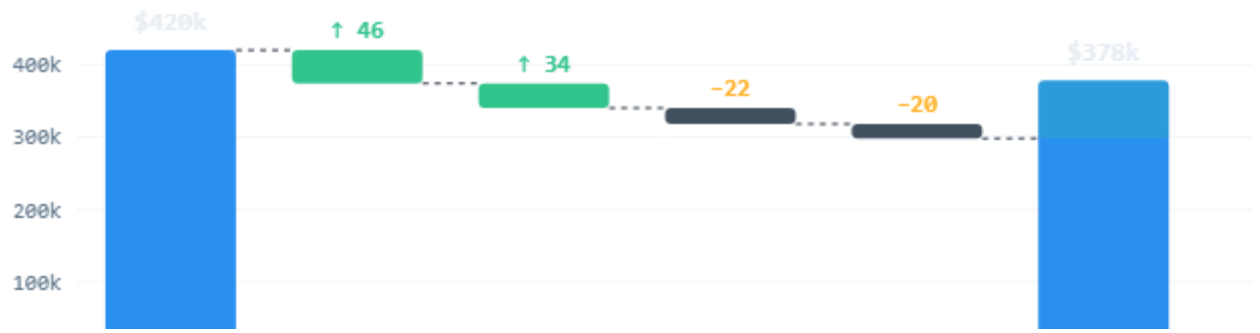
A line decision is a financial decision. The bridge from pieces-per-hour to dollars-per-year is where a simulation earns its keep.

OEE decomposes performance into Availability × Performance × Quality, so a single low number becomes a diagnosis. A **loss waterfall** then prices each cause — breakdowns, setups, scrap, imbalance — as a bar in dollars, turning "the line is slow" into "reducing ST-2 changeovers is worth $67k/year."

EXPLORABLE · LOSS-TO-CASH DECISION MODEL

From pieces lost to the financial case

The simulator prices each loss in $/year — select what to recover; the cards return the decision KPIs



✓ Breakdowns · +$46k/yr · capex $60k

✓ Setups / SMED · +$34k/yr · capex $25k

Scrap · +$22k/yr · capex $40k

Line balance · +$20k/yr · capex $55k

RECOVERED / YEAR	PAYBACK	ROIC VS WACC 12%	NPV · 5-YR @12%
$80k/yr	12.8 mo	94% vs 12%	$203k

Value-creating. ROIC beats the 12% cost of capital by 82 points and 5-year NPV is positive — the simulator turns "the line is slow" into a defensible investment case.

Nakajima (OEE); Goldratt (ToC); EVA = ROIC − WACC; NPV discounted at WACC

Lab 9 · Fig 9

---

10 / VIABILITY & HONEST SCOPE

Beat the cost of capital — and know where the tool stops

"It pays back" isn't enough. The real test is whether the change earns more than the money it ties up — and a tool you can trust is one that tells you where it stops.

Payback is intuitive but blind to time value. The real test of value creation is **ROIC above WACC**: a positive EVA spread means the change earns more than the capital it uses, every year — not just "pays back." And RASQI is a **flow-line simulator**: serial and assembly lines, built for accuracy with simplicity. It deliberately leaves out job-shop routing, re-entrant loops, and heavy AGV handling — supporting them would trade away the simplicity that lets anyone who knows the line run it. A model used outside its assumptions is worse than no model.

> **A simulator you can trust is one that tells you where it stops.**

Sources: Stern Stewart (EVA); Brealey, Myers & Allen, *Principles of Corporate Finance*; Law, verification & validation.

See these ideas on your own line

Every model in this handbook — the queue, the waterfall, the confidence interval — is one free month away from your real numbers.

▶ Try RASQI free — online

See what it models

BUILT IN THE TRADITION OF EXPLORABLE EXPLANATIONS — BRET VICTOR, BARTOSZ CIECHANOWSKI, DISTILL — WITH THE CHART DISCIPLINE OF THE FT VISUAL VOCABULARY & THE ECONOMIST. NO FRAMEWORKS, ONE FILE.