TECHNICAL WHITE PAPER

CourtRank

A Transparent Ranking System for Rotating-Partner Doubles Pickleball Ladders

Design rationale, algorithm specification, and Monte-Carlo validation

Prepared byVersion 1.0

Michael HofwellerJune 2026

Contents

Executive Summary

CourtRank is a ranking system designed for an eight-week, rotating-partner doubles pickleball ladder of roughly thirty-six players across six courts. It converts each player's DUPR rating into a starting position, then adjusts that position each week based on actual match results, re-sorting players into courts under a strict guardrail: **no player may move more than one court per week** in either direction.

The system is built on a margin-aware Elo engine with two additions tailored to this format. A **contribution split** handles the central difficulty of doubles ranking—how much credit each partner earns from a shared result—by combining a per-game heuristic with the natural experiment created by partner rotation. A **gated fast-track** lets mis-seeded newcomers reach their correct court quickly without destabilizing players who are already accurately placed.

Across 150 simulated seasons per scenario, roughly **nine in ten players finish within one court of their true skill level**, the guardrail was never violated, and rating inflation was negligible. The rest of this paper documents the reasoning, the mechanics, and the test evidence.

1. The Problem and Its Constraints

The league presents a deceptively simple request—rank the players each week—wrapped around constraints that interact in non-obvious ways. Three of them shaped every subsequent design decision.

1.1 Attribution in doubles is unidentifiable within a single game

From the final score of one doubles game, there is no statistical way to recover how much each partner contributed. The score is a single observation of a two-person system; any split between partners is an assumption, not a measurement.

The rotating-partner format, however, is a designed experiment. Each player partners several different people and faces shifting opposition, which makes each player the one common factor across their own four-plus games—and a player who wins regardless of partner reveals their strength through that consistency. The system therefore uses a defensible per-game heuristic and lets the rotation perform the real attribution across the session. In testing, a hidden strong player seeded identically to weaker peers separated from them by roughly 113 rating points after a single session, correctly identified in 84% of runs.

1.2 Stability and responsiveness pull in opposite directions

Capping how far a player can move each week—so that, for example, last place cannot leap to court two—directly limits how fast a mis-ranked player can climb to where they belong. The design must therefore let **ratings** move quickly while letting **ladder position** move slowly, and reconcile the two without letting a single dominant night rocket someone across the ladder.

1.3 A constraint from the rotation itself

With six players on a court, it is impossible for everyone to partner everyone exactly once: that would need fifteen pairings at two per game, or seven and a half games. The practical rotation is six games in which each player plays four, with four distinct partners. A five-player court is perfect—five games, every pair partnered once—which is why courts of five are used when attendance is not a multiple of six.

2. The CourtRank Algorithm

CourtRank is a margin-aware, contribution-weighted team Elo with a clamped ladder and one safety valve. It has five working parts.

2.1 Seeding

Every player's DUPR—real where available, an organizer-assigned shadow grade otherwise—is converted to a rating on a 1500-centered scale. Because the conversion is a straight line, sorting by seed rating gives the same order as sorting by DUPR: week-one courts are just the players listed highest to lowest and dealt six at a time. Each player is also tagged as carrying a verified or a shadow grade; this does not affect week-one placement but governs the fast-track in Section 2.4.

$$R_0 = 1500 + (\mathrm{DUPR} - 3.5) \times 200 \tag{1}$$

2.2 Per-game rating updates with margin

Each game is scored on its own merits. A team's rating is the average of its two partners,

$$R_{\mathrm{team}} = \frac{R_i + R_j}{2} \tag{2}$$

and the Elo expectation gives the probability that team A should beat team B given the rating gap between them:

$$E_A = \frac{1}{1 + 10^{(R_B - R_A)/400}} \tag{3}$$

The actual result blends a win/loss term with a point-margin term, weighted by β. With W the win indicator and s_f, s_a the points scored for and against,

$$S = \beta W + (1 - \beta)[\,0.5 + 0.5 \cdot \frac{s_f - s_a}{11}\,] \tag{4}$$

The team's rating change is then the usual Elo product, applied zero-sum across the two teams:

$$\Delta_{\mathrm{team}} = K\,(S - E) \tag{5}$$

A decisive 11–2 win therefore moves ratings roughly twice as far as an 11–9 result. Because games are scored against the ratings actually on the floor, the system is indifferent to which court a game is played on—players who cross between courts within a session need no special handling.

2.3 The contribution split

Within a team, a result is not divided evenly. Each partner's share of the swing is proportional to their rating, clipped so that no player ever absorbs more than 65% or less than 35%:

$$w_i = \mathrm{clip}(\frac{R_i}{R_i + R_j},\ 0.35,\ 0.65) \tag{6}$$

After the two shares are renormalized to sum to one, player i's rating change is

$$\Delta_i = 2\,w_i \cdot \Delta_{\text{team}} \tag{7}$$

so the higher-rated partner—the presumed carrier—gains more from a win and loses more from a loss. The clip keeps the per-game guess humble; the durable attribution comes from rotation, as a player who consistently lifts different partners accumulates credit no single game could justify.

2.4 Adaptive update size and the gated fast-track

New players use a larger update factor that decays toward the steady-state value over their first few games, where g is games played and K_0 = 56, K_∞ = 32:

$$K(g) = K_\infty + (K_0 - K_\infty)\,e^{-g/\tau_K} \tag{8}$$

A returning absentee receives a temporary multiplier on top of this, growing with the number of missed weeks a and capped so it cannot run away—a lightweight stand-in for formal uncertainty tracking:

$$m = \min(1.5,\quad 1 + 0.15\,a) \tag{9}$$

The most carefully tuned component handles uncertain players who **sweep a session**—win every game or lose every game. The system then computes the session performance rating P: the rating that would have made the day's results unsurprising, given the mean opponent rating $\langle O\rangle$, mean partner rating $\langle T\rangle$, and mean team score $\langle S\rangle$:

$$P = 2(\langle O\rangle + 400\log_{10}\frac{\langle S\rangle}{1-\langle S\rangle}) - \langle T\rangle \tag{10}$$

The rating is then pulled toward P by a strength that itself decays with experience (b the blend ceiling, g_p games played entering the session), capped at 0.85:

$$\lambda = \min(0.85,\quad b \cdot e^{-g_p/\tau_p} \cdot m) \tag{11}$$

A dead zone z suppresses the pull for gaps small enough to be session noise, and the move is clipped to $\pm c$:

$$R \leftarrow R + \lambda \cdot \mathrm{clip}(\max(0,\ |P-R| - z) \cdot \mathrm{sgn}(P-R),\ \ -c,\ c) \tag{12}$$

A clean sweep is roughly a six-percent fluke for a correctly seeded player but is near-certain for a gross mis-seed, which makes it a reliable trigger. Players holding a verified DUPR are never subject to this pull, which prevents the system from over-reacting to one unusual night from a player already well characterized.

2.5 Weekly re-rank and the guardrail

After each session every enrolled player is re-sorted by updated rating, and courts are dealt from that order. The governing rule is structural rather than statistical: no player's rank may change by more than six positions—one court—in a single week, in either direction. Writing r_{new} and r_{old} for a player's rank after and before the re-sort,

$$|\,r_{\text{new}}(p) - r_{\text{old}}(p)\,| \le 6 \quad \text{for every player } p \tag{13}$$

Ratings may change substantially in a night, but ladder position changes gradually, so no blowout carries a player from the bottom court to the top. Absent players keep their rating and position, frozen; new mid-season players enter at the rank implied by their seed and are treated as uncertain until they have played a few sessions.

3. Why Not an Off-the-Shelf System

Glicko-2 and OpenSkill model player uncertainty rigorously, but they are opaque to participants and more machinery than thirty-six players with a clean weekly structure require. Their best idea—let less-certain ratings move faster—is borrowed here through the adaptive update size and the uncertain-player flag. DUPR itself is proprietary and not built for week-to-week ladder placement. Plain Elo is the right backbone because every number it produces can be reconstructed in a spreadsheet—which matters the first time a player disputes a court assignment.

4. Simulation and Validation

The system was tested against a simulator in which thirty-six players hold hidden true skills, receive noisy real or shadow seeds, play out games rally by rally, and are subject to realistic weekly form variation. Four scenarios were each run for 150 simulated seasons. The headline measures are the rank correlation between true skill and rating, the fraction of players placed in their correct court, and the largest single-week upward move observed.

The rank correlation is reported as ρ (the Greek letter rho), the Spearman correlation between each player's true skill and the rating the system has assigned them. It asks a simple question: if you sort the players by the system's ratings, how closely does that order match their true order of ability? A value of $\rho = 1.0$ means the ordering is perfect, $\rho = 0$ means the ratings are no better than a random shuffle, and intermediate values fall in between—so $\rho = 0.90$ indicates the rankings are very nearly in true order, off only in the occasional close pair. It measures the order of players, not the exact rating values, which is what matters for assigning courts.

Scenario	Wk 1 ρ	Wk 8 ρ	Exact court	±1 court	Max up-move
A. Baseline (full attendance)	0.84	0.91	54%	92%	6
B. 12% weekly absences	0.84	0.90	54%	92%	6
C. Absences + two wk-3 joiners	0.84	0.88	51%	90%	6
D. Sandbagger (top-3, seeded last)	0.74	0.88	52%	90%	6

Table 1. Core results across four stress scenarios, 150 simulated seasons each. ρ is the Spearman correlation between true skill and rating.

By the end of a season, roughly 92% of players sit within one court of where their true skill belongs; absences and mid-season arrivals barely move that figure. Rating inflation under the rule that winners never lose points measured about a tenth of a point per player per season—negligible. Most importantly, across every run of every scenario, the largest single-week upward move was exactly six positions: the guardrail never leaked.

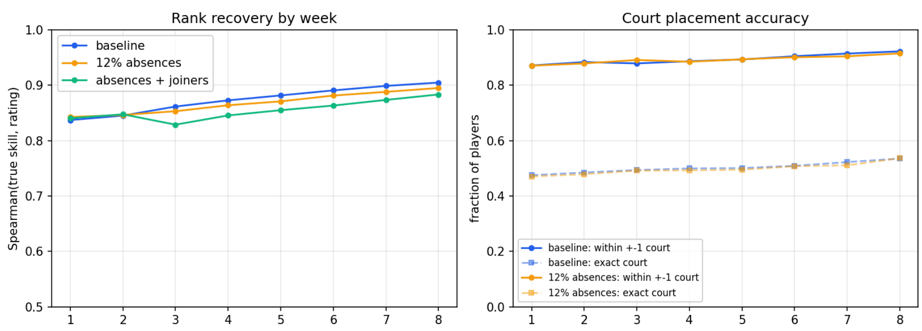

Figure 1. Rank-recovery (left) and court-placement accuracy (right) improve week over week and are robust to absences and joiners.

4.1 The hardest case: severe mis-seeds

The most demanding test drove the final design. A genuinely strong player seeded dead last initially climbed too slowly, because partner-diluted Elo updates are bounded by construction. An early fix that pulled every player toward their session performance rescued the sandbagger but hurt accuracy for the well-seeded majority. The resolution exploits a fact the organizer already knows—who holds a verified DUPR and who does not—by restricting the fast-track to shadow-graded players and joiners. This recovered full baseline accuracy while still letting a mis-seeded player climb steadily toward their true court.

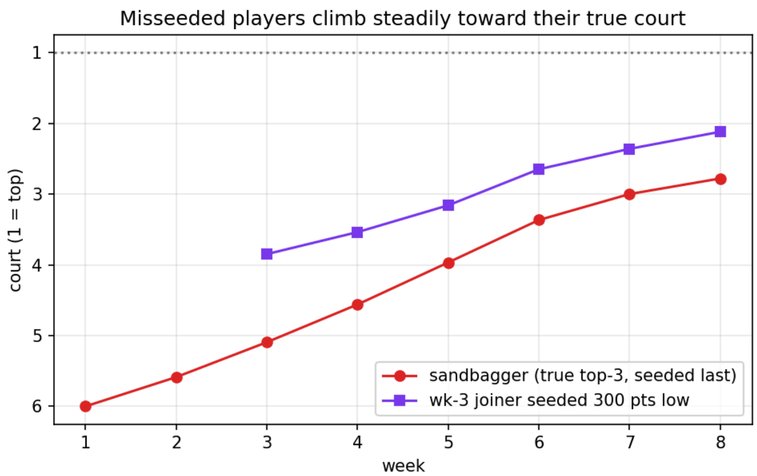

Figure 2. A last-seeded strong player and a mis-seeded mid-season joiner both climb steadily toward the top courts, one court at a time.

One honest limitation remains, and it follows from the guardrail rather than the engine. A severely mis-seeded player climbs steadily but may take most of the season to fully arrive, since movement is capped at one court per week. When an obvious grading error is visible in week one, the right remedy is for the organizer to amend that player's seed by hand rather than wait for the ladder to walk them up.

5. Operating the System

Week one. Sort every player by DUPR or shadow grade from highest to lowest and deal courts—top six to court one, and so on. Ties are common because DUPR moves in coarse steps; players at the same number are genuinely indistinguishable on day one, so list order is fair. Where a tie straddles a court boundary, breaking it in favor of the player with a real DUPR is a reasonable nicety but is not required.

Each week thereafter. Provide two things: the game scores from the session, or at minimum who won each game, and the list of who was present. The DUPR seeds are needed only once. By roughly the third week the original seeds have largely washed out and the ladder is driven almost entirely by results on the court.

Interventions. Correct a clearly erroneous seed by hand in the first week or two. Beyond that, the system is designed to run untouched, with the guardrail ensuring that no single week can produce an indefensible jump.

Appendix A. Recommended Parameters

The following settings are the validated defaults. They are exposed as configuration so the organizer can adjust them, but the values below reflect the full tuning described in Section 4.

Parameter	Setting and rationale
Seed conversion	1500 + (DUPR − 3.5) × 200; identical ordering to raw DUPR.
Update factor (K)	32 at steady state, 56 for a player's first games, decaying over roughly three weeks.
Margin weight	Half weight on win/loss, half on point margin.
Contribution clip	Each partner's share held between 35% and 65% of the team swing.
Fast-track	Fires only on a session sweep by an uncertain player; verified-DUPR players exempt.
Rank clamp	Six positions (one court) maximum movement per week, each direction.
Absence handling	Rating and rank frozen; temporary larger update factor on return.

Appendix B. Notation

Symbols used in the equations of Section 2. Angle brackets denote a session mean—an average taken over all of a player's games in a given week.

Symbol	Meaning
$R_i,\ R_{\text{team}}$	Rating of player i; team rating (the mean of the two partners).
E_A	Elo win expectation for team A against team B.
$S,\ W,\ \beta$	Blended game score in [0,1]; win indicator (1 win, 0 loss); margin weight.
$s_f,\ s_a$	Points scored for and against in the game.
$K(g)$	Update factor at g games played, decaying from K_0 = 56 to K_∞ = 32.
$\tau_K,\ \tau_p$	Decay constants (in games) for the update factor and the pull strength.
$\Delta_{\text{team}},\ \Delta_i$	Team rating change; player i's renormalized share of it.
w_i	Contribution share of player i, clipped to [0.35, 0.65].
$m,\ a$	Absence multiplier; number of consecutive missed weeks.
$\langle O\rangle,\ \langle T\rangle,\ \langle S\rangle$	Session-mean opponent rating, partner rating, and team score.
P	Session performance rating (the rating implied by the week's results).
$\lambda,\ b,\ z,\ c$	Pull strength; blend ceiling (0.6); dead zone; clip on the pull (±400).
$r_{\text{new}},\ r_{\text{old}}$	A player's rank after and before the weekly re-sort.

Appendix C. A Worked Example

This appendix carries one game and one short rivalry all the way through the arithmetic, so the formulas of Section 2 can be seen operating on real numbers. Every figure below is produced by the reference engine at the recommended parameters.

C.1 A single game, recorded and scored

Four players take the court. Their current ratings are A = 1580, B = 1500 (team 1) against C = 1540, D = 1520 (team 2). They play to 11 and team 1 wins 11–9. The organizer records only what happened on the court—who partnered whom, and the score:

Game	Team 1	Team 2	Score
1	A + B	C + D	11 – 9

Table C-1. The complete record an organizer enters for a game: partners on each side and the final score.

From that record the engine computes the update. The team ratings are the partner averages, R_1 = 1540 and R_2 = 1530, so team 1 is expected to win with probability E = 0.514 (equation 3). The 11–9 result gives a margin term of 0.591, which blends with the win to a score S = 0.795 (equation 4). The team rating swing is therefore $K(S - E)$ = 32 × 0.281 = +9.0 points to team 1, and the mirror image to team 2. The contribution split then divides each team's swing between its partners by rating (equations 6–7):

Player	Rating before	Change	Rating after
A	1580	+9.23	1589.23
B	1500	+8.76	1508.76
C	1540	−9.05	1530.95
D	1520	−8.94	1511.06

Table C-2. The same game after scoring. The swing is zero-sum, and within each team the higher-rated partner moves slightly more—A gains more than B, C loses more than D.

Notice the two properties the design promises. The winners' gains exactly offset the losers' losses, so the rating pool neither inflates nor deflates. And within team 1, the stronger player A gains 9.23 while B gains 8.76: the presumed carrier is credited a little more, the free-rider a little less, even though they shared the same win.

C.2 Two rivals over three weeks

A full thirty-six-player session is too large to reproduce by hand, so this example follows just two evenly matched mid-pack players—call them P and Q—who both begin at 1500, near the court-3 boundary. To keep the arithmetic legible, only two of each player's games per week are shown; the engine treats every game identically. Each line below is a recorded game and the resulting change to P or Q:

Week	Recorded game (team vs team, score)	P	Q
1	P+S beat M+N, 11–8	+17.08	—

Week	Recorded game (team vs team, score)	P	Q
1	Q+M beat P+T, 11–9	−11.75	+11.77
2	Q+S beat M+N, 11–7	—	+15.97
2	Q+T beat P+N, 11–9	−13.23	+13.50
3	P+T beat M+S, 11–9	+14.82	—
3	P+M beat Q+N, 11–8	+13.13	−13.72

Table C-3. Two games per player per week, with each player's rating change. A dash means that player was not in that game.

Tallying the changes week by week shows the rivalry move on the ladder. Both start level; P's strong opening week is offset by a head-to-head loss to Q, while Q's week 2 surges ahead; P then recovers in week 3, including a head-to-head win that pulls the two back together:

	Start	End wk 1	End wk 2	End wk 3
P rating	1500	1505	1492	1520
Q rating	1500	1512	1541	1528
P court	3	3	4	3
Q court	3	3	2	3

Table C-4. The same rivalry as ratings and the court each would be assigned. Court bands here are illustrative for this small example.

Two things are worth drawing out. First, even a clear week—Q jumping to court 2 after week 2—moves a player by a single court at a time, never more, which is the guardrail at work. Second, the system is responsive without being twitchy: a strong or weak night shifts a rating by 10–17 points, enough to change a court over a few weeks but not enough for one lucky game to upend the ladder. By week 3 the two rivals, having traded the lead, sit back within a few points of each other—an honest reflection of two players who really are about the same standard.

This document accompanies the reference implementation (courtrank.py) and its test harness (run_tests.py), which reproduce every figure and table herein.