

Acceleration in the NBA: Towards an Algorithmic Taxonomy of Basketball Plays

Philip Maymin
NYU-Polytechnic Institute
Brooklyn, NY 11201
Email: philip@maymin.com

Abstract

I filter the 25-frames-per-second STATS/SportVu optical tracking data of 233 regular and post season 2011-2012 NBA games for half-court situations that begin when the last player crosses half-court and end when possession changes, resulting in a universe of more than 30,000 basketball plays, or about 130 per game. To categorize the plays algorithmically, I describe the requirements a suitable dynamic language must have to be both more concise and more precise than standard X's and O's chalk diagrams. The language specifies for each player their initial starting spots, trajectories, and timing, with iteration as needed. A key component is acceleration. To determine optimal starting spots, I compute burst locations on the court where players tend to accelerate or decelerate more than usual. Cluster analysis on those burst points compared to all points reveals a difference in which areas of the court see more intense action. The primary burst clusters appear to be the paint, the top of the key, and the extended elbow and wing area. I document the most frequently accelerating players, positions, and teams, as well as the likelihoods of acceleration and co-acceleration during a set play and other components intended to collectively lead to an algorithmic taxonomy.

1 Introduction

Basketball coaches preach and teach execution but objectively measuring execution, let alone estimating the contribution of execution on winning, has eluded analysis. Part of the problem is the language describing the desired execution. Basketball plays are routinely drawn up on chalkboards with standard static graphical notation, but the precise timing is often explained only orally to the huddled players. Here I introduce a dynamic algorithmic approach to concisely encode theoretical basketball plays and I describe its key characteristics. Important inputs to the language are the frequencies and locations of player acceleration.

I analyze optical tracking data on the 25-frames-per-second positional data of 233 regular season and post season 2011-2012 NBA games for half-court situations that begin when the last player crosses the half-court line and end when the offense no longer has possession. This subset, which by construction excludes both fast and secondary breaks, is ideal for analyzing set plays.

I document the teams, positions, and players in the dataset who exhibit the most and least frequent acceleration both on offense and on defense. Further, I evaluate the incidence of co-acceleration when multiple players experience bursts nearly simultaneously. I also determine the primary locations of bursts, compare and contrast them with the primary locations among inertial states, and evaluate the optimal number of such clusters. In addition, I provide new graphic tools, both static and dynamic, to ease analysis of these important issues. Taken together, these results build a roadmap towards an algorithmic taxonomy of basketball set plays.

Past research on optical tracking data in basketball includes [1], [2], and [3]. Acceleration in basketball has been studied from a physiological perspective, c.f. [4] in which the authors found that college basketball players were superior in terms of acceleration to non-athletes, but to my knowledge this is the first work exploring acceleration from optical data and its implications for an algorithmic approach to categorizing halfcourt set plays.

2 Data

The three-dimensional SportVu optical tracking data from STATS LLC assigns to each player on the court an ordered (x, y) pair representing the position of their center of mass on a regulation 94 feet by 50 feet NBA court, and assigns to the ball an additional z coordinate specifying its height above the ground. These coordinates are recorded 25 times per second.

In addition, event identification information is automatically assigned to frames satisfying certain criteria such as dribbles, field goal attempts, and the like.

The data set covers 233 regular-season and post-season games during the lockout-shortened 2011-2012 season. Due to differential adoption, these games are skewed towards the teams that installed the required technology. Figure 1 shows the breakdown of home games for which the data was available.

Figure 1: Home Games in Sample, by Team

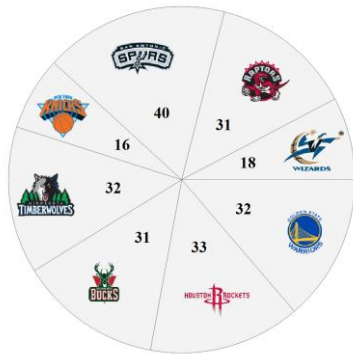
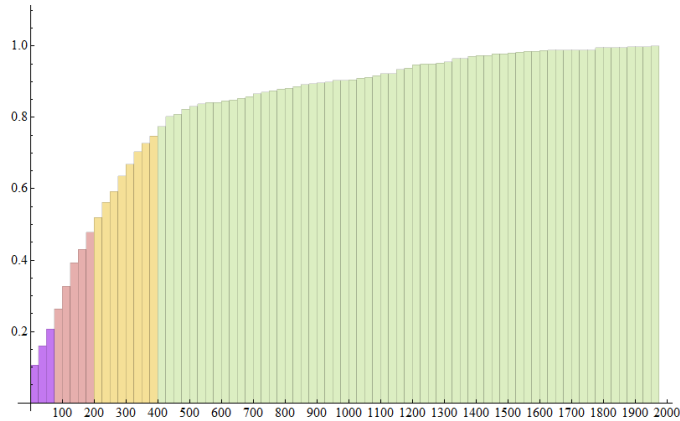


Figure 2: Cumulative Histogram of the Number of Plays per Player



I filter the sequences of these coordinates within games to create subsequences of halfcourt set plays, defined and implemented as follows. A halfcourt set play begins when all ten players are on the same side of the court, the opposite side from the previously recorded set play, and the ball is inbound, within 20 feet of the halfcourt line, and nearest to one of the offensive players. The set play ends when the difference in time from the previous snapshot is more than 1/25 of a second (for example, a timeout has been called or the quarter has ended), when any player appears in the back court, or when the ball is nearest to a defensive player.

This definition is intended to capture choreographed set plays rather than improvised fast breaks or secondary breaks; in other words, possessions where each player's movements are the result of intentional practice. To focus on halfcourt set plays, out-of-bounds set plays are excluded, except for situations where the ball is thrown back inbounds to near halfcourt, at which point the definition above obtains and a presumptive regular set play can run.

The data set reduces to a universe of 30,950 plays lasting on average 180 frames each, or about 7 seconds. The data set includes location information for 10 players as well as the basketball, comprising more than 60 million coordinates in total. Some of the 456 distinct players are involved in more plays than others, because their team has more home games in the sample or because they have more playing time. Figure 2 shows the cumulative histogram of the number of plays each player is involved with in our universe. About three quarters of the players participate in at least one hundred distinct plays.

3 Methodology

Acceleration is computed as the second difference of the Euclidean distances between sequential moving average positions of a player on the court, divided by the standard gravity $g = 32.174 \text{ ft/s}^2$, multiplied by $625 = 25^2$ because the frames in the optical data are 1/25 of a second apart, and multiplied by 10 to express the acceleration in units of deci-g's, where $1 \text{ dg} = 0.1 \text{ g}$ is one-tenth of the standard acceleration of gravity g .

The window for the moving average, used to smooth the data for better accuracy in measuring the acceleration, is five. Finally, the most extreme accelerations are clipped because they are likely the result of random noise.

Specifically, the first smoothed position in a play for a particular player is calculated as the average of raw position numbers one through five for the player; the second position is the average of raw position numbers two through six; and so on. The acceleration is computed as the second difference, so the first acceleration is the second difference between the third and the first smoothed positions. Thus, it is the second difference between the average of raw positions three through seven and one through five.

In short, the computation for a single acceleration number requires eight frames from the optical data. While this may seem like a lot, it actually represents only $8 * 40 \text{ ms} = 320 \text{ ms}$, which is literally in the blink of an eye: the experimentally measured blink duration is $334 \pm 67 \text{ ms}$ [5].

Bursts of high acceleration are rare. The table at the right of Figure 3 shows the conditional and unconditional frequency of occurrence as a function of the amount of acceleration. Acceleration is rare, and becomes rarer still for greater acceleration. Note that computed accelerations in excess of 7 dg, comprising less than one percent of the total, were clipped to 7 dg.

Figure 3 shows the frequency of acceleration on offense by team, with teams sorted in increasing order by their inertial proportion. The width of the bars represents the amount of data available for that team in our sample. The Houston Rockets spent the least time accelerating; the Boston Celtics the most. The corresponding team graph for defensive acceleration by team (not shown) has two substantive differences: the Milwaukee Bucks were even more inertial on defense than on offense, while the San Antonio Spurs moved up from the bottom into the middle of the pack.

Figure 3: Frequency of Offensive Acceleration by Team
(Bar widths correspond to amount of data available in sample)

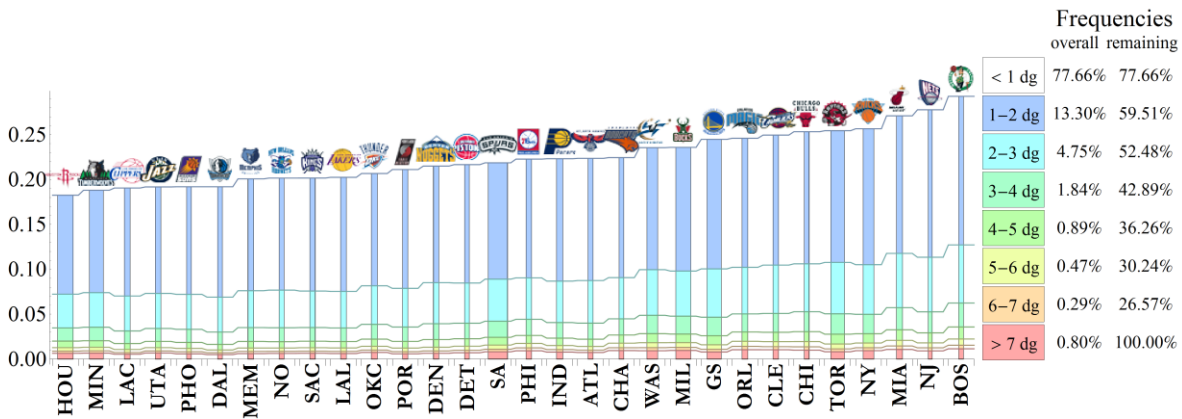


Table 1: Most and Least Frequent Accelerators per Position (Percent of Time > 1 dg)

Frequent Accelerators	Infrequent Accelerators
Solomon Alabi (41.98%) Tyson Chandler (35.05%)	Marcus Camby (18.46%) Marc Gasol (22.39%)
Joel Anthony (36.90%) Andrea Bargnani (34.90%)	Troy Murphy (17.81%) Tyrus Thomas (18.57%)
Kris Humphries (32.41%) Taj Gibson (31.12%)	Marcus Morris (12.05%) Matt Bonner (14.74%)
Carmelo Anthony (34.34%) LeBron James (32.27%)	Bobby Simmons (12.02%) Chandler Parsons (12.96%)
Paul Pierce (29.06%) Kevin Durant (26.20%)	Chase Budinger (11.02%) Danilo Gallinari (15.69%)
J.R. Smith (30.03%) Gary Forbes (27.94%)	Courtney Lee (10.43%) Danny Green (13.45%)
Kobe Bryant (26.82%) Ray Allen (25.56%)	Shannon Brown (13.64%) Daequan Cook (14.84%)
Toney Douglas (28.60%) Dwyane Wade (27.72%)	Gary Neal (14.38%) Daniel Gibson (14.55%)
Deron Williams (33.36%) Baron Davis (26.85%)	Derek Fisher (12.52%) Kyle Lowry (13.54%)

Figure 4: Frequency of Offensive Acceleration by Position

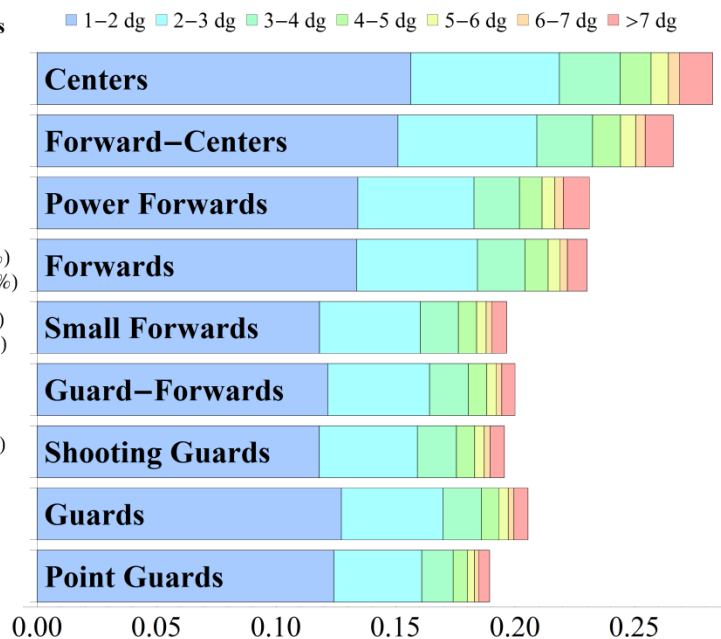
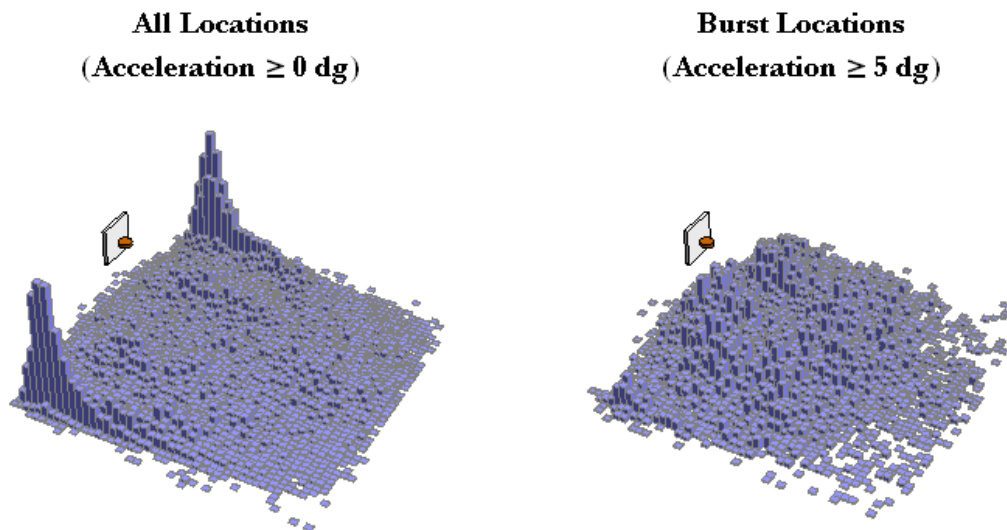


Figure 4 shows the frequency of acceleration on offense by position, with positions sorted by average height of players at that position. Centers have more bursts than guards, partially because they are more likely to set picks. Even among forwards and centers, bigger players tend to exhibit more extreme accelerations. Not shown are accelerations less than 1 dg in absolute magnitude; thus, point guards are not accelerating, i.e. inertial, 81 percent of the time while centers are inertial only 72 percent of the time. Note that inertial players are not necessarily standstill, they merely have constant velocity. The corresponding graph for defensive acceleration by position (not shown) does not substantially differ.

Table 1 lists the players with the highest and the lowest frequency of accelerating at least 1 dg while on offense (in other words, those with the lowest frequency are the ones with the highest inertia). The table would look somewhat different for different thresholds of acceleration, e.g. if restricting only to accelerations greater than 3 dg instead of 1 dg, but the top 10 names in each category tend to be relatively stable. Further, essentially the same names appear on the corresponding defensive table (not shown): players seem to accelerate, or not, based on who they are, not based on whom they are guarding.

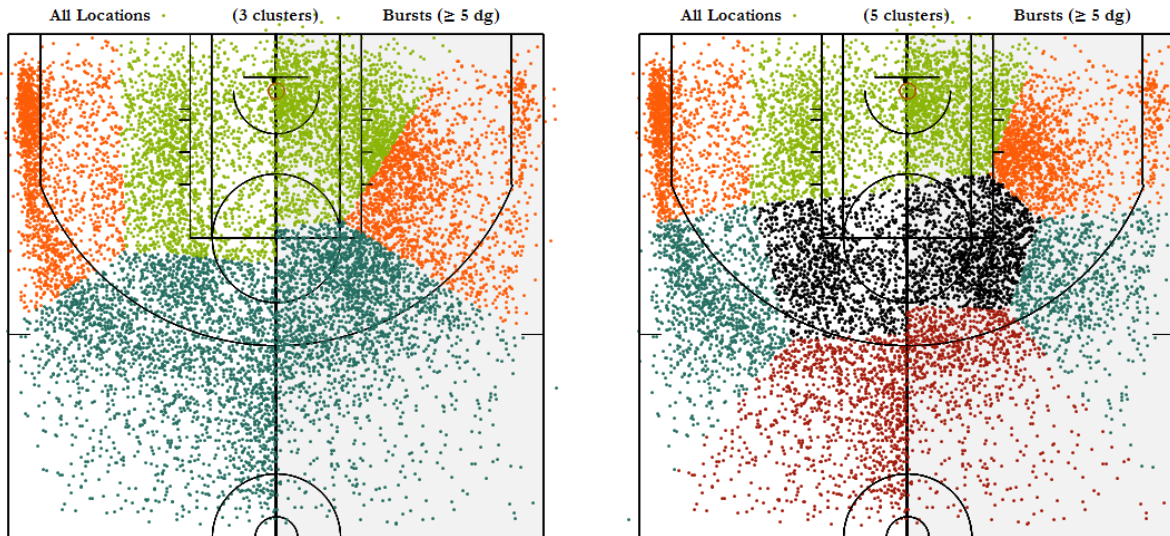
The spatial distribution of all players differs markedly from accelerating players. Figure 5 displays the three-dimensional histogram of halfcourt player locations for a random subsample of all players, and for a random subsample of players at times when they are accelerating or decelerating at a magnitude of 5 dg or greater. Random subsamples were used to facilitate display: the entire universe of halfcourt set plays contains nearly 30 million offensive player positions.

Figure 3: Histogram of Positions of All Players and Accelerating Players



A cluster analysis further highlights the differences while also suggesting common burst areas. Figure 6 shows these results. Note the differences between the left-hand sides and right-hand sides of each halfcourt. As with Figure 5, the corner three is a popular location, albeit not one with much acceleration. Among burst points, the three primary clusters appear to be the paint, the top of the key, and to a lesser extent, the combined area of the extended elbows and wings. Extending to five clusters separates out the elbow and the corner wings as additional areas, but do not appear to be as well demarcated as the three clusters. Thus, we are justified in treating the three cluster graph as an appropriate model for burst points, which will become starting points in our language.

Figure 4: Cluster Analysis of Positions of All Players and Accelerating Players



4 Play Language and Specification

Bursts of extreme acceleration tend to happen in the paint, at the top of the key, or in the extended elbow and wing areas. With each area on the left and on the right of the court, there are six possible starting and ending spots for player trajectories:

- 1) **LP** and **RP**: Left Paint and Right Paint
- 2) **LK** and **RK**: Left Key and Right Key
- 3) **LW** and **RW**: Left Wing and Right Wing

In principle, player trajectories may happen simultaneously or after previous trajectories are finished. Based on the standard “X’s and O’s” graphical notation that does not specify an order for trajectories, it indeed appears as if all movements happen simultaneously. In practice, we can explore how often burst points happen at the same time.

How often do multiple players accelerate simultaneously? If by “simultaneously” we mean the exact same frame (1/25 of a second), then the answer is virtually never. But if by “simultaneously” we mean “within one second of each other,” then we can count the number of times within all rolling 25-frame subperiods that no players, exactly 1 player, exactly 2 players, exactly 3 players, exactly 4 players, or all 5 players were accelerating.

Figure 5: Proportions of Co-Accelerating Players

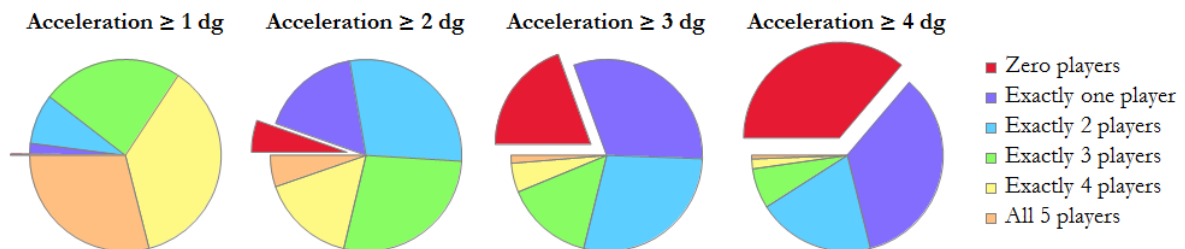


Figure 7 shows the pie charts of these counts for acceleration thresholds of 1 dg, 2 dg, 3 dg, and 4 dg. The likelihood of multiple players accelerating or decelerating at a magnitude of at least 1 dg in any given one-second interval is nearly 100 percent: in other words, mild acceleration is the norm for most players most of the time. The story changes at more extreme bursts, however. The co-acceleration likelihood drops to about 75 percent for 2 dg, 50 percent for 3 dg, and 25 percent for 4 dg.

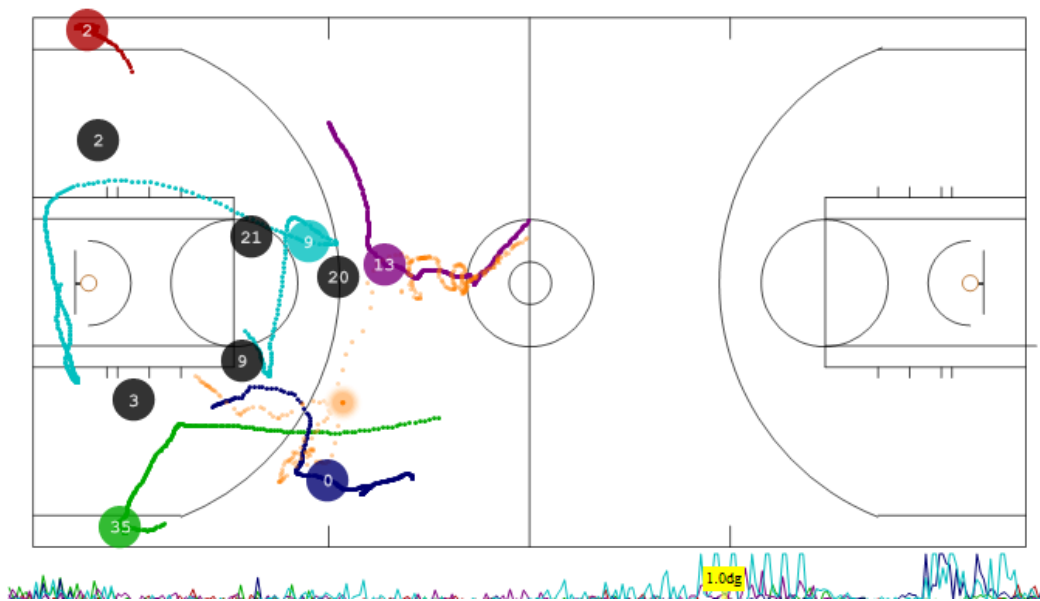
It may be useful and simplifying to assume that no player begins a new trajectory while another player is in the middle of his trajectory. This is a relatively non-constricting assumption because trajectories may be broken up into parts, e.g. a player crossing the court from one wing to the other may, if necessary, be modeled as running first to the nearest painted area, then to the other, and then finally to the opposite wing.

Therefore, the state of a play at any given time requires noting for each player which burst cluster they are currently in and which one they are aiming towards. If they are not accelerating, then they may be said to aiming to the same cluster they are currently in. Finally, the basketball itself needs to be modeled as well; because we also need to know only its current location and destination, it can be treated as a sixth player on the court.

An example may help illustrate the approach. Figure 8 shows the snapshot from a video examination of the final assisted field goal of the Thunder against the Spurs in Game 6 of the 2012 Western Conference Finals. The play started with a little under two minutes remaining as James Harden (#13) dribbled over halfcourt, defended by Manu Ginobili (#20). The snapshot in Figure 8 occurs about two-thirds of the way through the play, at a time after Harden has passed the ball to Russell Westbrook (#0) but before Westbrook has caught it.

Note the five individual offensive trajectories. Thabo Sefolosha (#2) exhibits essentially no acceleration or velocity in RW. Kevin Durant (#35) accelerates from LK to LP then to LW. Harden accelerates from RK to RW though he doesn't quite reach it before the play ends. (However his acceleration is nevertheless important as it helps draw Ginobili away from the action.) Serge Ibaka (#9) has the most complicated route and the most amount of screens, accelerating from LP to RP, then to RK, then finally to LK, where his final pick frees up Westbrook to take and make the jumper and extend the Thunder's lead to four points, essentially sealing their victory in the game and the series. Note also that Westbrook's acceleration towards the end of the play coincides with Ibaka's; similarly, Harden's acceleration starting at the snapshot shown also coincides with Ibaka's.

Figure 6: 2012 Western Conference Finals, Game 6, OKC@SA, Q4 1:49 – 1:36



5 Conclusions

Calculating and analyzing the accelerations of each offensive and defensive player in each halfcourt set from optical tracking data, along with novel static and dynamic visualizations, helps shed light on the rare but critical bursts that help define basketball plays and suggests a road towards an algorithmic description and ultimately a taxonomy of plays. Future research directions may include attempting to predict future accelerations from past data alone, implementation of a tool to help advance scouts categorize plays, and exploring the relationships between acceleration and co-acceleration on field goal percentage, spacing, and execution variability.

Acknowledgments

I am grateful to Ken Catanella, David Griffin, Sam Hinkie, Brian Kopp, Allan Maymin, Daryl Morey, Eugene Shen, and Ryan Warkins for discussions around the concepts addressed in this paper, and to STATS LLC for their data.

References

[1] Sandy Weil, "The Importance of Being Open: What Optical Tracking Data Can Say About NBA Field Goal Shooting," MIT Sloan Sports Analytics Conference, presentation at <http://sloansportsconference.com/?p=661>, Mar. 2011.

[2] Allan Maymin, et al., "Individual Factors of Successful Free Throw Shooting," *Journal of Quantitative Analysis in Sports*, vol. 8, no. 3, Oct. 2012.

[3] Rajiv Maheswaran, et al., "Deconstructing the Rebound with Optical Tracking Data," MIT Sloan Sports Analytics Conference Proceedings, Mar. 2012.

[4] Gillam G. McKenzie, "Identification of anthropometric and physiological characteristics relative to participation in college basketball," *National Strength & Conditioning Association Journal*, vol. 7, no. 3, pp. 34-36, Jun. 1985.

[5] Frans VanderWerf, et al., "Eyelid Movements: Behavioral Studies of Blinking in Humans Under Different Stimulus Conditions," *Journal of Neurophysiology*, vol. 89, no. 5, pp. 2784-2796, May 2003.