CINEWAVE: AN AFFORDABLE 6DOF ROBOTIC ARM FOR CINEMA ROBOTICS

A Thesis Submitted in Partial Fulfillment of the
Requirements of the Renée Crown University Honors Program at
Syracuse University

RIO HARPER

Candidate for Bachelors of Science in Visual Communications
and Renée Crown University Honors
4/30/2026

Honors Thesis Faculty Advisor: Milton Santiago

Honors Thesis Reader: Zhenyu Gan

Abstract

This project presents the design and development of a custom 6-degree-of-freedom robotic arm optimized for precision cinematography applications. The system addresses the challenge of achieving cinema-grade camera positioning while maintaining cost-effectiveness compared to commercial solutions employing harmonic drives throughout. Through analysis of pixel-level motion requirements for a Sony FX3 at 5-meter subject distances, the project establishes joint-specific backlash tolerances ranging from 6 arcminutes at the shoulder to 30 arcminutes at the wrist, with corresponding design torques of 70 Nm for the shoulder down to 3 Nm for the wrist assembly.

The methodology combines harmonic drive reducers for the high-torque base joints (J1-J3) with an innovative cable-driven capstan system using Dyneema DM20 for the wrist joints (J4-J6), achieving the required backlash performance through antagonistic cable pairs and proper tensioning rather than expensive precision gearing. Structural components are manufactured via high-strength 3D printing using PLA-CF and pultruded carbon fiber tubes, in which the filament enables complex integrated geometries (such as cable channels and hollow shafts) that would be cost-prohibitive to machine from aluminum. The capstan drive based wrist design reduces cost and weight significantly comparedpared to traditional timing belt or harmonic drive approach.. This hybrid approach demonstrates that professional-grade cinematography robotics can be achieved at significantly reduced cost by strategically matching transmission technologies and employing recent developments in engineering grade 3D printed filament to achieve previously impossible technology using traditional CNC methods.

Executive Summary

Modern audiences expect dynamic, immersive camera movement in film and television. Watch the opening sequence of Apple TV's *Severance* Season 2, and you will see a camera sweeping around an actor with inhuman smoothness and precision, creating an unsettling atmosphere that would be impossible to achieve by hand. This kind of shot requires a robotic camera arm, a machine that can position a cinema camera in three-dimensional space with millimeter-level accuracy while executing perfectly repeatable movements.

The problem is access. Productions that use systems like the BOLT, the industry-standard robotic camera arm, are typically budgeted at ten million dollars or more. The equipment itself costs upwards of $200,000, requires a specialized crew to operate, and is simply not available to independent filmmakers, student productions, or small studios. When I worked as a Director of Photography on a university capstone horror film, I wanted to incorporate the kind of controlled camera movement that defines the genre's most effective sequences. Our university equipment shop offered nothing that could accomplish this. The most accessible mainstream alternative, a handheld gimbal like the DJI Ronin, only stabilizes the camera in three axes and still relies entirely on a human operator to move through space. It cannot reach, extend, or sweep around a subject the way a robotic arm can, and it inherits all the natural inconsistency of human motion.

This project aims to bridge that gap by designing and building a six-degree-of-freedom robotic camera arm with professional-grade precision at a target material cost of $2,500.

A six-degree-of-freedom robotic arm can position its end (in this case, a camera) anywhere within its reach while also orienting it in any direction. Think of how your own arm works: your

shoulder, elbow, and wrist combine to let you place your hand almost anywhere around your body and point it in any direction. This robot replicates that capability with six motorized joints, each controlled by software to coordinate smooth, precise motion.

The arm is designed around a specific set of requirements derived from real cinematography needs. At a typical filming distance of five meters with a standard lens, a single pixel on a 1080p display corresponds to about 1.7 millimeters of physical movement. Any mechanical "slop" or looseness in the joints, known as backlash, must be small enough that it does not produce visible jitter or drift in the final image. Through geometric analysis, I determined that the joints closest to the camera can tolerate more looseness than those at the base, because the same angular error at the base translates to much larger displacement at the end of a one-meter arm.

This insight drives the core design decision of the project: using different technologies for different joints based on their specific requirements. The three joints at the base and elbow, which demand the tightest precision, use devices called harmonic drives. These are compact, high-precision gear systems that can reduce motor speed while multiplying torque with almost no backlash. They are also expensive, at roughly $300 each compared to only around $120 for the wrist joints. The three joints at the wrist, which can tolerate more looseness, use a cable-driven system instead. By wrapping high-strength synthetic rope around precision drums in opposing pairs, the cables can be kept under constant tension, eliminating slack without the cost and weight of additional harmonic drives. This hybrid approach delivers professional-level performance where it matters most while keeping the overall system affordable and lightweight.

The project follows a systematic engineering design process. I began by establishing quantitative requirements: How much weight must the arm support? How precise must each joint be to avoid visible errors on screen? How far must it reach? These questions have concrete, calculable answers based on the physics of camera optics and mechanical systems.

From these requirements, I evaluated candidate technologies for each joint, comparing cost, weight, precision, and complexity. The hybrid harmonic-drive and cable-drive architecture emerged from this analysis as the best balance of performance and affordability. As it stands on March 13th, 2026 I have completed the CAD modeling for the entire arm, and have built and tested the wrist (J4-J6) wrist, with the rest of the arm being completed in the next month.

This project addresses a real and underserved need in independent filmmaking and educational settings. A functional, affordable robotic camera arm would give student filmmakers and indie productions access to visual techniques currently reserved for major studio projects. Beyond individual creative empowerment, I envision this as a tool for interdisciplinary collaboration: an engineering student could build the arm for a film program, bridging technical and artistic disciplines in exactly the kind of partnership that universities should foster.

More broadly, this project demonstrates an approach to precision robotics that may have applications beyond cinematography. Any field requiring accurate, repeatable positioning, from scientific instrumentation to automated inspection, could benefit from the cost-reduction strategies explored here. For my own development as an engineer, this work synthesizes my background in humanoid robotics and cinematography with the demands of high-precision mechanical systems, building skills directly applicable to professional robotics careers.

The camera has always been the audience's eye into a story. This project aims to give more storytellers control over how that eye moves.

Table Of Contents

Chapter One: Background

The camera in an expert's hands is more than just a way to record what’s in front of them, it is the audience's surrogate body within the world of a film. How that body moves through space fundamentally shapes the viewer's emotional experience. Research in cognitive film theory, particularly work from Tallinn University's Baltic Film, Media and Arts School, has demonstrated that moving cameras create measurably higher viewer immersion than static shots [1]. This effect is explained through the theory of embodied cognition: when viewers watch camera movement onscreen, their sensorimotor systems reconstruct analogous motion mentally, creating a visceral sense of participation in the scene rather than passive observation.

Different types of camera motion evoke distinct psychological responses. A slow dolly push toward a character's face builds tension and intimacy. A rapid tracking shot alongside a running subject creates urgency and kinetic energy. A smooth arc around a figure can convey contemplation, power, or unease depending on context. These are not merely stylistic flourishes but fundamental tools of visual storytelling, as essential to a director's vocabulary as dialogue or performance.

The horror genre, in particular, relies heavily on camera motion to generate unease. Consider the opening sequence of Apple TV's Severance Season 2, in which the camera sweeps around the protagonist with inhuman smoothness and impossible precision. The movement itself communicates something deeply wrong about the world the character inhabits. This effect cannot be achieved with a handheld camera, which introduces the subtle tremors and corrections of a human operator. It cannot be achieved with a gimbal alone, which stabilizes orientation but still depends on a person to translate the camera through space. It requires robotic motion: perfectly repeatable, precisely controlled, and decoupled from human physiology. The forces against the protagonist are deeply mechanized, planned and fluid, and always one step ahead.

A BOLT robotic arm being used to create the running sequence in *Severance* (2025)

The distinction matters because audiences perceive the difference, even if they cannot articulate it. A camera moved by human hands carries an implicit point of view, a physical presence behind the lens. A camera moved by robotic control can become disembodied, omniscient, or alien. This quality is why productions ranging from Marvel films to high-end commercials invest in robotic camera systems when a shot demands movements that transcend human capability.

Existing Solutions and Their Limitations

Professional Robotic Systems

At the high end sit purpose-built cinema robots manufactured by companies like Mark Roberts Motion Control (MRMC), a Nikon subsidiary that has become the industry standard for robotic cinematography. Their flagship product, the BOLT, is a six-axis robotic arm capable of moving at speeds up to 2 meters per second on its base and 12 meters per second when combined with a linear track. The arm can carry payloads up to 20 kilograms and achieve positional repeatability of ±0.02 millimeters [2], meaning it can execute the same complex movement hundreds of times with imperceptible variation between takes.

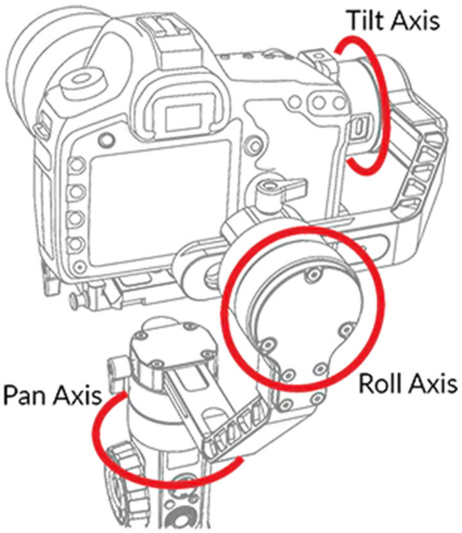
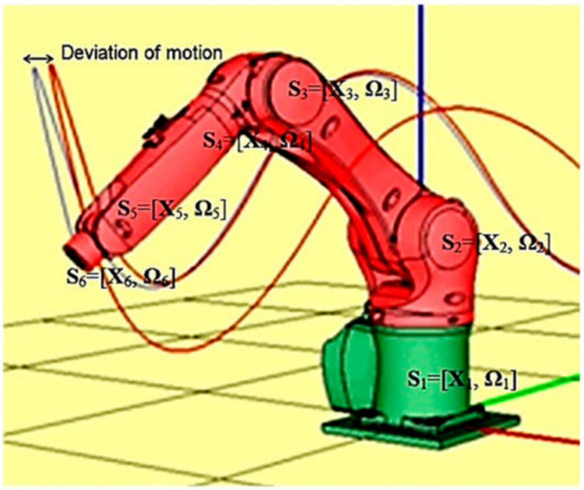
The BOLT and its variants (the longer-reach BOLT X, the more compact BOLT Jr+) dominate high-end production. They appear in Marvel films, Apple TV productions, and national advertising campaigns. However, their cost places them far beyond independent production budgets. The entry-level BOLT Mini Model Mover, designed primarily for product photography and small-object work, starts at approximately $70,000. The full-size BOLT and Milo systems cost substantially more and typically require rental rather than purchase.

Beyond equipment cost, these systems demand specialized expertise. MRMC's Flair motion control software has a steep learning curve, and operating the robot safely around talent requires trained technicians. A BOLT shoot typically involves a dedicated motion control operator in addition to the standard camera department. This combination of capital expense and crew overhead means that robotic camera motion remains effectively inaccessible outside productions with budgets in the millions.

Handheld Gimbals and Stabilizers

At the accessible end of the market sit motorized gimbals like the DJI Ronin series. The Ronin 2 can stabilize cameras weighing up to 13.6 kilograms (30 pounds) across three axes: pan, tilt, and roll. It achieves 0.02 degrees of angular precision and can maintain stable footage even when mounted to vehicles traveling at 75 miles per hour [3].

However, a gimbal is not a robotic arm. It stabilizes the camera's orientation but provides no capability to translate the camera through space. When a camera operator carries a Ronin through a scene, the smoothness of the resulting footage depends entirely on their physical skill at walking, turning, and controlling their own body. A gimbal cannot reach out and sweep around a subject. It cannot extend from floor level to overhead in a single continuous motion. It cannot repeat a complex three-dimensional movement with the consistency needed for visual effects compositing.

 A visualization of DoF in a DJI Ronin	 A visualization of DoF in a 6 DoF Robotic Arm

More fundamentally, gimbals inherit all the natural variation of human motion. This is sometimes desirable, as when a documentary aesthetic calls for footage that feels observational and present. But when a shot requires the mechanical precision that signals something beyond human perception, a gimbal cannot deliver it.

Open-Source and DIY Robotic Arms

The maker and educational robotics community has produced numerous open-source robotic arm designs, including projects like Thor, Arctos, PAROL6, and the Annin AR4. These designs have made six-degree-of-freedom robotics accessible to hobbyists and students, with complete builds possible for as little as a few hundred dollars in parts.

However, these arms are designed for educational purposes or light industrial tasks, not cinema applications. Thor, for example, stands 625 millimeters tall and can lift payloads of only 750 grams [4]. Arctos, while more sophisticated, remains a desktop-scale device. None of these projects prioritize the specific combination of reach, payload, and precision that cinematography demands.

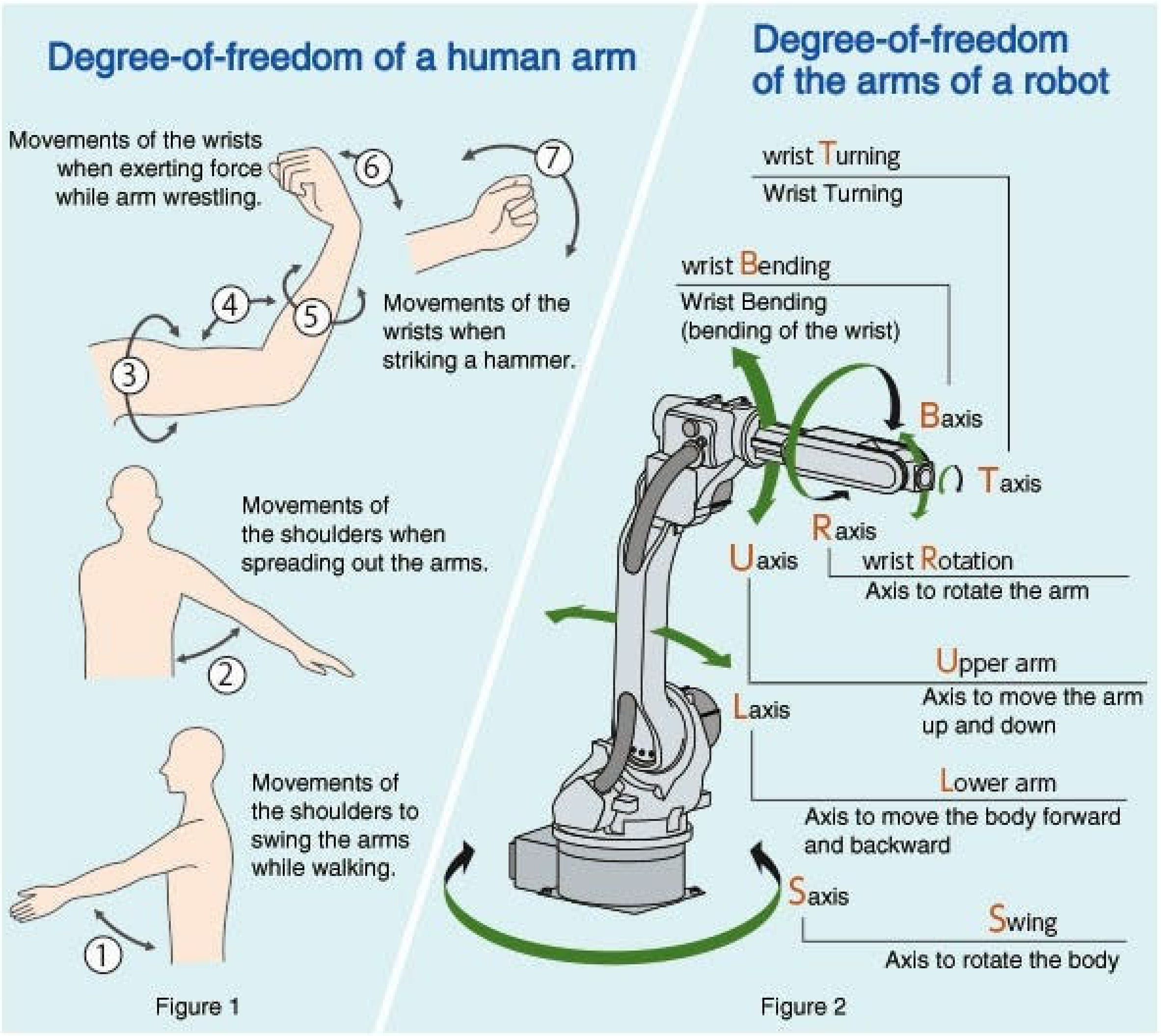
The gap is not merely one of scale. Open-source robot arms typically use 3D-printed plastic gears or timing belt reductions that accept significant backlash as a reasonable tradeoff for low cost. For a pick-and-place robot learning task or an educational demonstration, a few degrees of slop in the joints is acceptable. For a camera that must hold rock-steady while transitioning between precise positions, that same backlash creates visible jitter and renders footage unusable.

This thesis project exists precisely because this gap is addressable. The same advances in accessible manufacturing that enable open-source robotics, particularly high-strength 3D printing, can be applied to a design that takes cinematography requirements seriously from the outset.

Key Concept/Terminology

Degrees of Freedom

A robotic arm's degrees of freedom (DOF) describe how many independent ways it can move. A simple hinge has one degree of freedom: it can rotate around a single axis. A human arm, from shoulder to fingertips, has approximately seven degrees of freedom. To position an object arbitrarily in three-dimensional space (specifying both its location and its orientation), a minimum of six degrees of freedom is required. This is why most industrial manipulators and cinema robots use a six-DOF configuration.

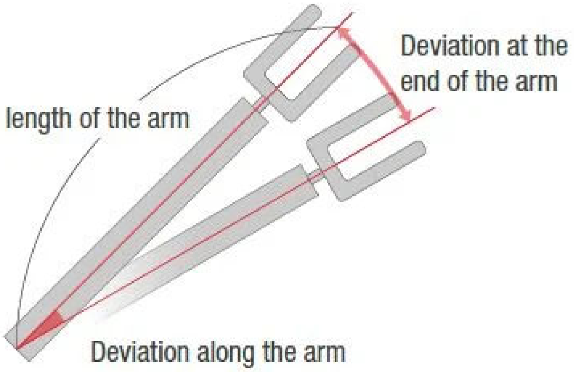


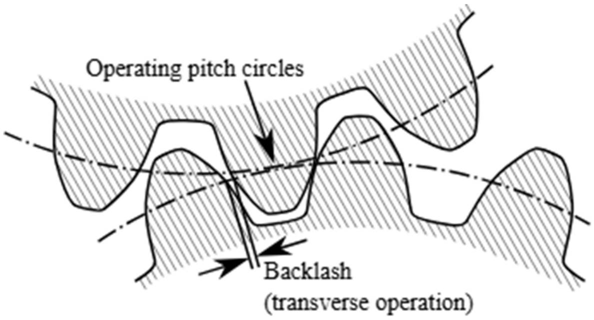
An illustration of robotic arms compared to a human arm [5]

The specific arrangement of those six joints matters. This project uses a configuration common in industrial robotics: the first three joints (base, shoulder, elbow) position the wrist in space, while the final three joints (often called the wrist pitch, yaw, and roll) orient the end effector (the camera). This separation simplifies both the mechanical design and the mathematics of motion control.

Backlash

Backlash refers to the looseness or play in a mechanical system. When a motor reverses direction, backlash is the small amount of "dead zone" during which the motor shaft rotates but the output does not move. In geared systems, it arises from the necessary clearance between meshing teeth. In belt drives, it can come from belt stretch or improper tensioning.





For a camera robot, backlash is the primary enemy of image quality. Even a fraction of a degree of angular backlash at a joint can translate to millimeters of camera displacement at the end of a one-meter arm. If that displacement exceeds what corresponds to a multi pixel movement at the subject distance, the audience will perceive jitter. The backlash requirements derived in this project's specifications section emerge directly from this pixel-level analysis.

Gear Reduction and Torque Multiplication

Electric motors typically spin fast but produce relatively little torque. Moving a heavy camera slowly and precisely requires the opposite: low speed and high torque. Gear reduction accomplishes this tradeoff by reducing output speed while proportionally increasing output torque.

A harmonic drive (also called a strain wave gear) is a specialized reduction mechanism prized in robotics for its high ratio, compact size, and extremely low backlash. The mechanism uses a flexible element that deforms slightly as it engages with a rigid outer ring, creating smooth motion with essentially zero play (AKA backlash) [6]. Harmonic drives are standard in industrial robots and cinema systems, but they are expensive, with even small units costing several hundred dollars.

Cable and Capstan Drives

An alternative to geared reduction is the cable drive, in which a synthetic rope wraps around a drum (capstan) attached to the output shaft while the motor pulls on the cable. By using antagonistic pairs of cables, one pulling in each direction, the system can be kept under constant tension, eliminating the slack that would otherwise introduce backlash. Cable drives are lighter and less expensive than harmonic drives, though they require careful tensioning and cannot achieve the same extreme reduction ratios in a compact package. Additionally, the design of the drum inherently limits the max rotation of the joint.

Chapter Two: Design Requirements

This section derives the quantitative specifications that drive the mechanical design. Rather than adopting arbitrary performance targets, each requirement emerges from first principles: what precision is necessary for the camera motion to be invisible to the audience, what payload must the arm support to accommodate professional equipment, what reach enables useful cinematographic framing, and what torques result from these choices.

Payload and Reach Specifications

Payload Specification

The target payload of 1.5 kilograms represents a deliberate positioning within the landscape of camera equipment.

What 1.5 kg enables: A Sony FX3, the reference camera for this project, weighs approximately 715 grams with battery and memory card installed. Modern compact cine lenses average below 700 grams. A typical configuration pairing the FX3 with a 50mm lens therefore weighs approximately 1.4 to 1.5 kg, fitting comfortably within the design envelope [7]. This combination delivers full-frame 4K video with professional color science and manual focus control suitable for narrative filmmaking.

Why not lighter: Reducing the payload target below 1 kg would push the system into territory already served by existing open-source robot arms. Projects like Thor (750 g payload) and similar educational designs can handle smartphones and action cameras adequately. However, these platforms cannot support the larger sensors, interchangeable lenses, and professional features that distinguish cinema-quality footage from consumer video. A lighter payload specification would solve an already-solved problem.

Why not heavier: The Sony FX6, a popular compact cinema camera one tier above the FX3, weighs approximately 890 grams body-only but reaches 2.59 kg in a typical shooting configuration with viewfinder, grip, battery, and lens. Supporting this class of camera would require a payload capacity of 3 to 4 kg, and this modest increase in payload has severe consequences for the mechanical system.

The critical issue is rotational inertia. Unlike a weight held close to a pivot point, a payload at the end of a 1-meter arm creates enormous resistance to angular acceleration. Rotational inertia scales with mass but also with the square of the distance from the axis of rotation:

$$I = m \cdot r^2$$

where I is the moment of inertia, m is the mass, and r is the distance from the rotation axis.

For the base joint J1, rotating the entire arm assembly, even a modest payload increase at the distal end produces a disproportionate increase in the torque required to accelerate and decelerate the system. Doubling the payload from 1.5 kg to 3 kg does not just double the base joint torque; it approximately doubles the dynamic torque contribution while the moment arm remains constant. The real challenge is that higher payloads demand stiffer structures to maintain precision, which adds mass throughout the arm, further increasing inertia at every upstream joint creating a feedback loop of engineering complexity that I don’t have the timetable or funds for.

This is why the MRMC BOLT systems capable of carrying 20+ kg payloads cost tens of thousands of dollars and weigh hundreds of kilograms, while achieving the same basic function of moving a camera through space.

Reach Specification

The arm's total reach of 1 meter, measured from the base rotation axis to the camera mounting point, enables the fundamental cinematographic movements this system is designed to perform.

A one-meter reach allows the camera to sweep in an arc around a seated or standing subject while maintaining useful framing. For a medium close-up of a person's face at approximately 1 meter subject distance, the arm can traverse nearly 180 degrees of arc while keeping the subject in frame. This enables the "orbit" shots characteristic of robotic cinematography, where the camera circles a subject with mechanical smoothness.

The reach also permits vertical travel from near floor level to above head height when the arm is configured appropriately on a tripod, supporting the dramatic reveal shots that move from low angle to high angle in a single continuous motion.

Increasing reach beyond one meter would provide additional creative flexibility but would substantially increase the torque requirements at the shoulder and elbow joints as explained in the payload specification section above. Because torque equals force times moment arm, each additional 10 cm of reach increases the shoulder torque requirement by roughly 10% for the same payload. A 1.5-meter arm supporting 1.5 kg would require shoulder torques comparable to a 1-meter arm supporting 2.25 kg.

Deriving Backlash Tolerances from Pixel-Level Analysis

The core insight driving this project's precision requirements is that backlash becomes visible to the audience when it causes camera displacement exceeding one pixel at the subject distance. Any mechanical looseness smaller than this threshold is functionally invisible; any looseness larger produces perceptible jitter or drift.

Step 1: Calculate the Horizontal Field of View

The horizontal field of view (HFoV) determines how much of the scene the camera captures. For a full-frame sensor with a standard lens:

$$\text{HFoV} = 2 \cdot \arctan\left(\frac{w_{\text{sensor}}}{2 \cdot f}\right)$$

where *w*_sensor is the sensor width and *f* is the focal length.

The Sony FX3 has a sensor width of 35.9 mm. Using a 50mm lens (a standard focal length for narrative work):

$$\text{HFoV} = 2 \cdot \arctan\left(\frac{35.9 \text{ mm}}{2 \cdot 50 \text{ mm}}\right) = 2 \cdot \arctan(0.359) = 2 \cdot 19.75^\circ = 39.5^\circ$$

Step 2: Calculate Scene Width at Subject Distance

For a typical subject distance of 5 meters (appropriate for medium and wide shots where robotic arm movements are most visible):

$$w_{\text{scene}} = 2 \cdot d \cdot \tan\left(\frac{\text{HFoV}}{2}\right)$$

$$w_{\text{scene}} = 2 \cdot 5\ \text{m} \cdot \tan(19.75^\circ) = 10\ \text{m} \cdot 0.359 = 3.59\ \text{m}$$

Step 3: Calculate Linear Displacement per Pixel

For 1080p video (1920 horizontal pixels), the scene width maps to the pixel count:

$$\Delta x_{\text{pixel}} = \frac{w_{\text{scene}}}{\text{horizontal pixels}} = \frac{3.59\ \text{m}}{1920} = 1.87\ \text{mm}$$

This value, approximately 1.9 mm, represents the maximum allowable camera displacement before the error becomes visible as a full pixel shift. This is our backlash budget.

Step 4: Convert Linear Displacement to Angular Tolerance by Joint

The critical insight is that the same angular backlash at different joints produces different linear displacements at the camera, depending on the moment arm (distance from joint to camera).

For a joint at distance *L* from the camera, an angular backlash *θ* produces linear displacement:

$$\Delta x = L \cdot \theta$$

Rearranging to find the maximum allowable backlash angle:

$$\theta_{\text{max}} = \frac{\Delta x_{\text{pixel}}}{L}$$

Applying this formula to each joint:

J1/J2 (Base and Shoulder): These joints are approximately 1.0 m from the camera (full arm length).

$$\theta_{\max} = \frac{1.87\ \text{mm}}{1000\ \text{mm}} = 0.00187\ \text{rad} = 0.107^\circ = 6.4\ \text{arcmin}$$

J3 (Elbow): Located approximately 0.5 m from the camera (half arm length).

$$\theta_{\max} = \frac{1.87\ \text{mm}}{500\ \text{mm}} = 0.00374\ \text{rad} = 0.214^\circ = 12.9\ \text{arcmin}$$

J4/J5/J6 (Wrist joints): Located approximately 0.15 to 0.20 m from the camera sensor.

$$\theta_{\max} = \frac{1.87\ \text{mm}}{175\ \text{mm}} = 0.0107\ \text{rad} = 0.613^\circ = 36.8\ \text{arcmin}$$

Summary of Backlash Requirements

Joint	Distance to Camera	Max Backlash (degrees)	Max Backlash (arcmin)
J1 (Base)	1.0 m	< 0.11\degree	< 6.5 arcmin
J2 (Shoulder)	1.0 m	< 0.11\degree	< 6.5 arcmin
J3 (Elbow)	0.5 m	< 0.21\degree	< 13 arcmin
J4 (Wrist Pitch)	~0.20 m	< 0.5\degree	< 30 arcmin
J5 (Wrist Yaw)	~0.18 m	< 0.6\degree	< 35 arcmin
J6 (Wrist Roll)	~0.15 m	< 0.7\degree	< 42 arcmin

These values explain why the design uses different transmission technologies for different joints. The base and shoulder joints demand backlash below 10 arcminutes, which effectively requires harmonic drives or similarly precise mechanisms. The wrist joints can tolerate 30+ arcminutes, making properly tensioned cable drives a viable and far less expensive alternative.

For design margin, the project adopts conservative round-number specifications: less than 6 arcminutes for J1/J2, less than 12 arcminutes for J3, and less than 30 arcminutes for J4/J5/J6.

Torque Requirements by Joint

With payload, reach, and configuration established, the torque requirements for each joint can be calculated. These calculations consider static loads (holding the arm and camera stationary against gravity) and provide a foundation for motor and gearbox selection.

Static Torque Analysis

Static torque at each joint equals the gravitational force on all masses distal to that joint, multiplied by the horizontal distance from the joint axis to each mass's center of gravity. The worst case occurs when the arm is fully extended horizontally.

Assumptions:

- Camera payload: 1.5 kg at the end effector
- Link 2 (forearm): estimated 1.0 kg, center of mass at 0.25 m from elbow
- Link 1 (upper arm): estimated 1.5 kg, center of mass at 0.25 m from shoulder
- Wrist assembly: estimated 0.5 kg at 0.9 m from shoulder

J3 (Elbow): Supports the forearm, wrist assembly, and payload.

$$\tau_{J3} = g \cdot \left(m_{\text{payload}} \cdot L_{\text{forearm}} + m_{\text{wrist}} \cdot L_{\text{forearm}} + m_{\text{forearm}} \cdot \frac{L_{\text{forearm}}}{2} \right)$$

$$\tau_{J3} = 9.81 \cdot (1.5 \cdot 0.5 + 0.5 \cdot 0.45 + 1.0 \cdot 0.25)$$

$$\tau_{J3} = 9.81 \cdot (0.75 + 0.225 + 0.25) = 9.81 \cdot 1.225 = 12.0 \text{ Nm}$$

J2 (Shoulder): Supports everything distal to the shoulder.

$$\tau_{J2} = g \cdot (m_{\text{payload}} \cdot L_{\text{total}} + m_{\text{wrist}} \cdot 0.9 + m_{\text{forearm}} \cdot 0.75 + m_{\text{upper arm}} \cdot 0.25)$$

$$\tau_{J2} = 9.81 \cdot (1.5 \cdot 1.0 + 0.5 \cdot 0.9 + 1.0 \cdot 0.75 + 1.5 \cdot 0.25)$$

$$\tau_{J2} = 9.81 \cdot (1.5 + 0.45 + 0.75 + 0.375) = 9.81 \cdot 3.075 = 30.2 \text{ Nm}$$

J4/J5/J6 (Wrist): Support only the camera payload at a short moment arm (approximately 0.15 m from the wrist center to the camera's center of gravity).

$$\tau_{\text{wrist}} = g \cdot m_{\text{payload}} \cdot L_{\text{CoG}} = 9.81 \cdot 1.5 \cdot 0.15 = 2.2 \text{ Nm}$$

Design Torque with Safety Factor

Applying a 2.0x safety factor to account for dynamic loads during acceleration, manufacturing tolerances, and long-term reliability:

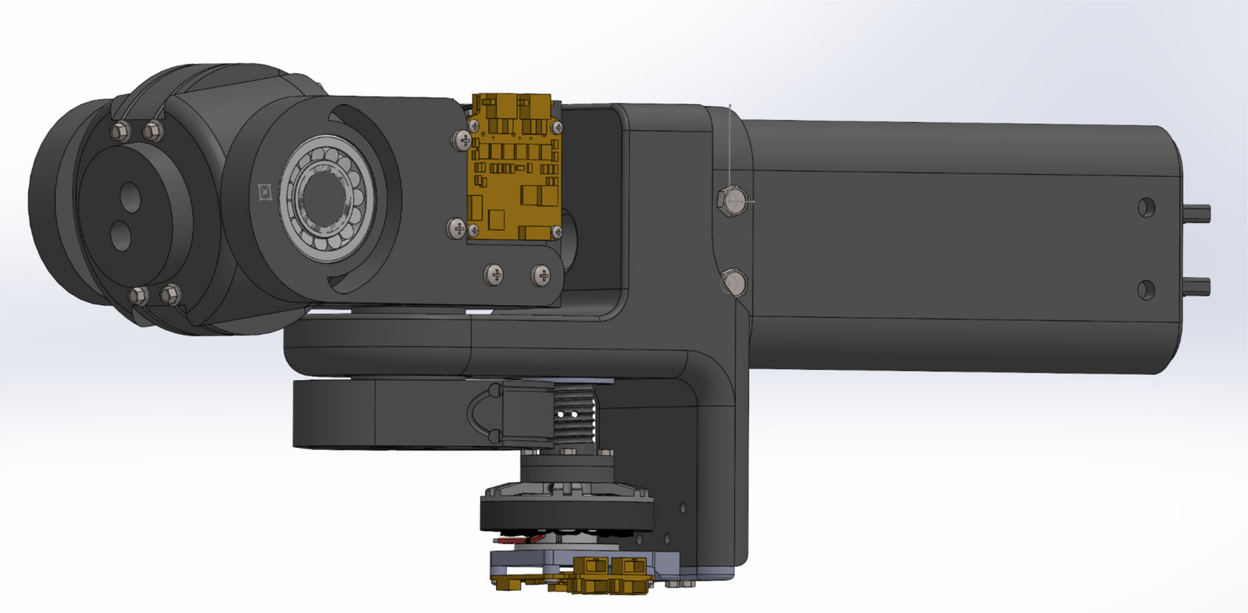
Joint	Calculated Static Torque	Design Torque (2.0x)
J1 (Base)	~15 Nm	~30 Nm
J2 (Shoulder)	~30 Nm	~60 Nm
J3 (Elbow)	~12 Nm	~25 Nm

J4/J5 (Wrist)	~2.2 Nm	~4.5 Nm
J6 (Roll)	~0.8 Nm	~1.6 Nm

Chapter Three: System Architecture

The Hybrid Drive Strategy: Why different joints need different solutions

The preceding analysis reveals a fundamental asymmetry in the requirements across the kinematic chain. The base joints demand tight backlash tolerances and must transmit substantial torques, while the wrist joints operate under relaxed precision requirements but are critically sensitive to mass. A uniform approach (using harmonic drives throughout, for example) would either over-engineer the wrist (adding unnecessary cost and weight) or under-engineer the base (compromising image quality). The hybrid strategy adopted here matches transmission technology to the specific demands of each joint group, optimizing the system as a whole rather than any individual component.



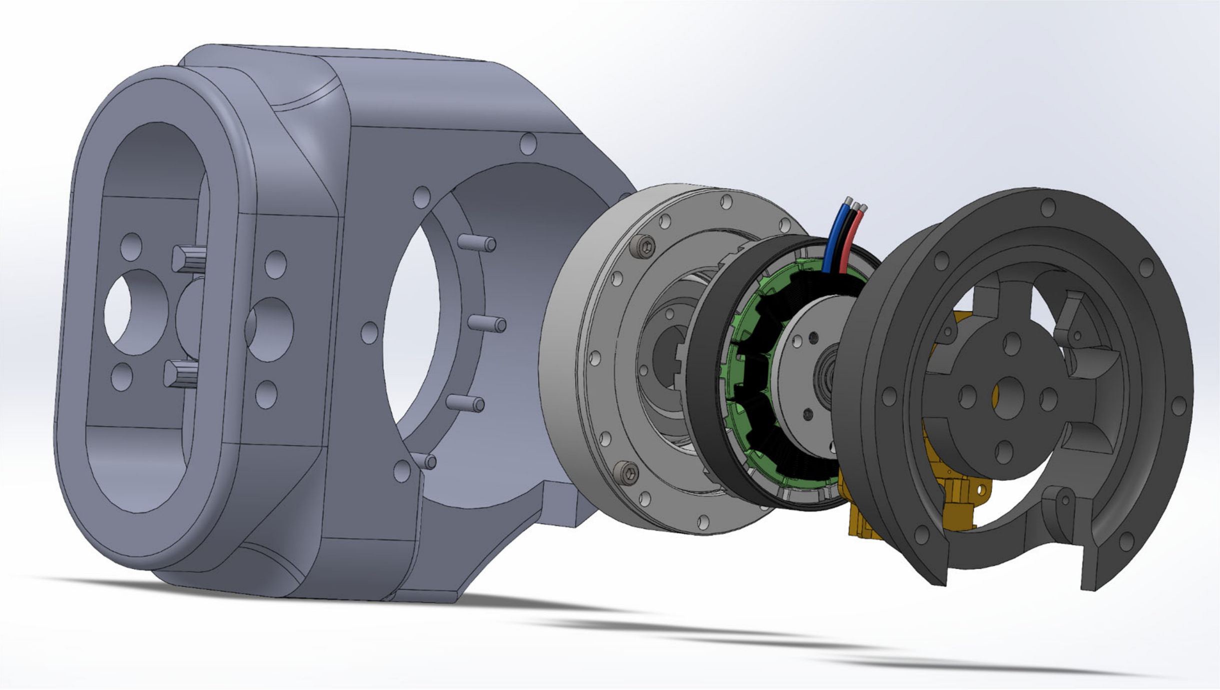
J1-J3: Harmonic Drives

The base rotation (J1), shoulder pitch (J2), and elbow pitch (J3) joints share a common challenge: any angular error at these positions propagates through the full length of the arm to the camera. As derived in the backlash analysis, J1 and J2 require sub-6-arcminute precision, while J3 requires sub-12-arcminute precision. These tolerances effectively mandate harmonic drive technology.

A harmonic drive (strain wave gear) achieves near-zero backlash through its operating principle. Unlike conventional gears where teeth mesh with clearance, a harmonic drive uses a flexible spline that deforms elastically to engage a rigid circular spline. The continuous elastic engagement eliminates the discrete tooth-to-tooth clearances that produce backlash in conventional gearing. Commercial harmonic drives routinely achieve backlash specifications below 1 arcminute, providing substantial margin against the derived requirements.

The torque multiplication these joints require further favors harmonic drives. The shoulder joint must deliver approximately 60 Nm of design torque, yet the motors selected for this project (MJBots mj5208) produce only 1.7Nm at their peak. An 80:1 reduction ratio transforms this motor torque into the 136 Nm theoretical output capacity needed to provide comfortable margin above the 60 Nm requirement. Harmonic drives achieve such ratios in a single, compact stage, whereas achieving equivalent reduction with spur gears or timing belts would require multiple stages, each introducing additional backlash and compliance.

The cost of this precision is, quite literally, cost. At approximately $300 per unit, the three harmonic drives represent nearly $900 of the project's $2,500 budget; over a third of the total allocated for a system that includes motors, controllers, structural materials, and electronics. This expense is justified only because these joints genuinely require the precision that harmonic drives provide. Extending this approach to the wrist joints would add another $900 while delivering precision far exceeding what the wrist requires, violating the project's cost-effectiveness goals without improving image quality.



J3 Harmonic Drive + motor in a 3D printed housing

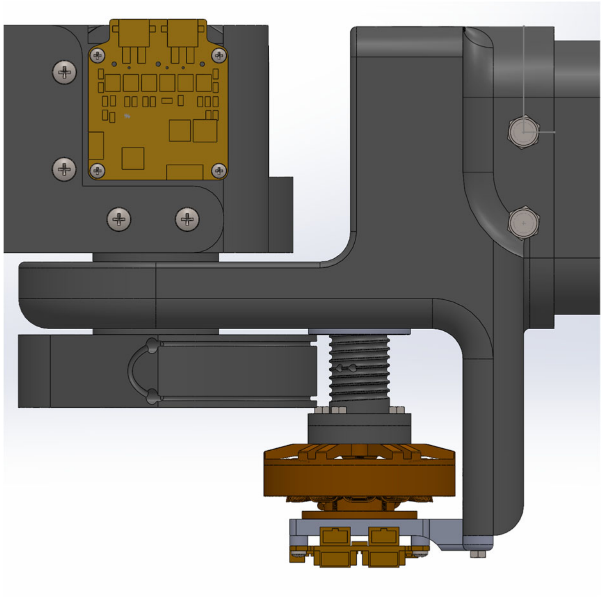
J4-J5: Capstan Drives

The first two degrees of freedom for the wrist employ a cable-driven capstan drive in a serial-based design, with both the driving pulley and driven sector drive being 3D printed. Notably, the J5 (pitch) uses the J6 housing as the capstan itself, giving it a signature quasi-spherical design. This skips an unneeded mechanism that would extend the bulk at the end of the arm, and makes the kinematic path cleaner as a result.

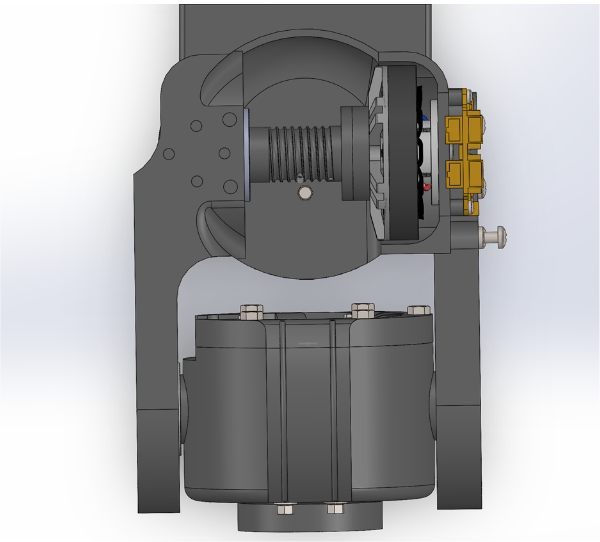
Both capstan drives use the same diameter of pulley of 20mm which had a good balance of strength from the torisonal forces being applied (too thin and it risks snapping under load) but still small enough to create a significant speed reduction with the driven sector drive without giant unwieldy diameters.

Each capstan stage provides a 5:1 mechanical advantage through multiple cable wraps around precision-ground drums. The antagonistic tensioning inherent to the dual-cable configuration eliminates backlash entirely. The cables are always in tension, so there is no dead band when reversing direction. This exceeds the 30 arcminute requirement by a substantial margin, providing headroom for any cable stretch or settling that may occur over the system's operational life. This design uses 1mm diameter Dynemma DM20 (a ultra-high molecular weight polyethylene) tendon, which has incredibly low elasticity which reduces the need for constant retensioning, a common disadvantage of cable driven systems [8].

Notably, J5 uses the J6 housing as the driven sector, which removes a traditional interminate sector drive and makes the kinematic path simpler, and creates the unique look of the J6 housing



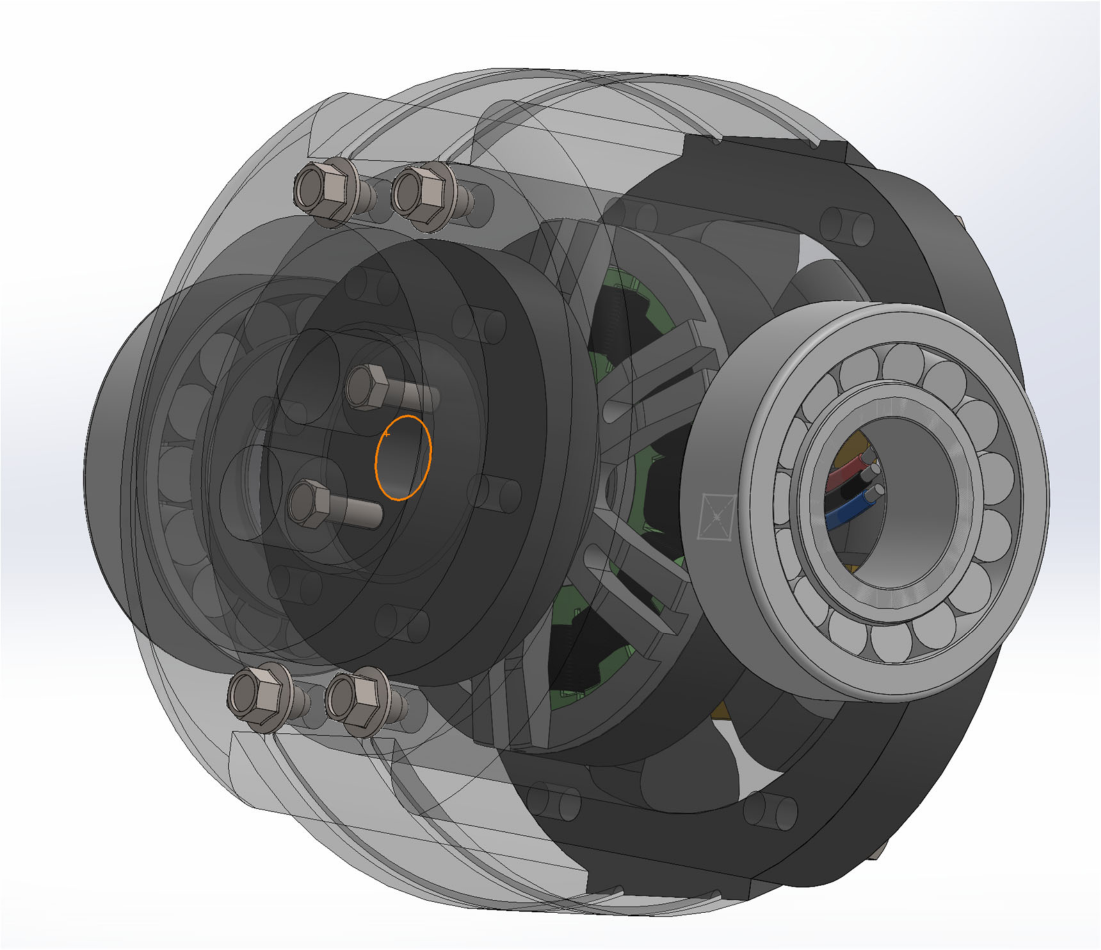
J4 (yaw) Capstan Design



J5 (pitch) Capstan Design with J6 housing being the sector

J6: Direct Drive

Due to J6s low torque requirements (~1.6Nm), this allows for the motor to drive J6 directly without a speed reducer. Thereby, it’s directly attached to a RU42 crossed roller bearing that is the end effector of the arm.



J6 Direct Drive + Housing

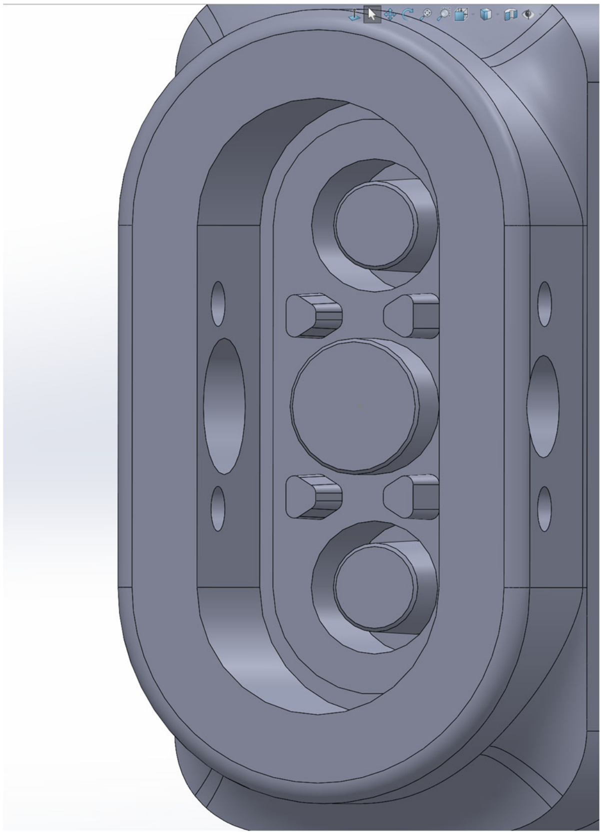
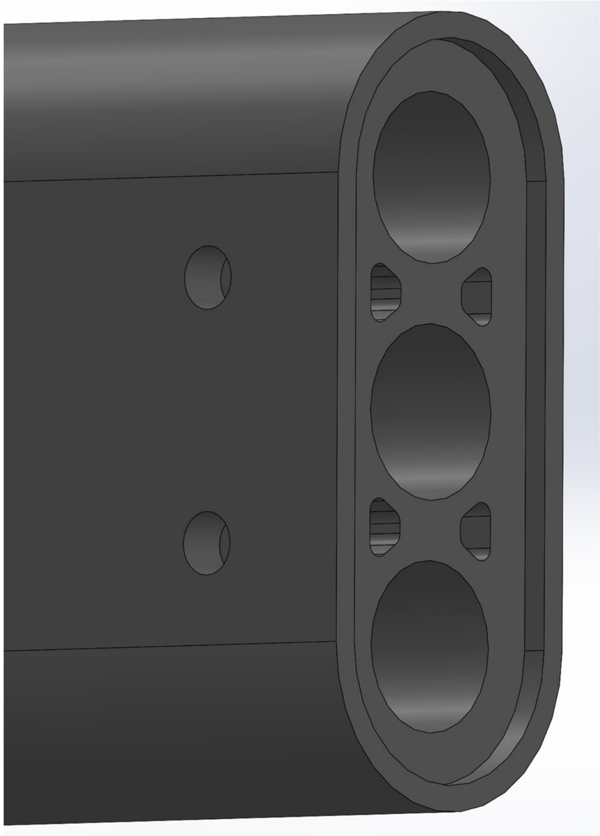
Chapter Four: Base and Arm Joints (J1-J3)

Structural Design of Links 1 and 2

Knowing that no matter the relative strength of my 3D printed filament choice, it is ultimately going to be inferior to non additive options. Thereby, I opted to use two pultrated parallel carbon fiber tubes in the arm links that run from one joint, through the arm link into the other joint housing. Pultrated carbon fiber, whose largest U.S. manufacturer resides in Syracuse (represent!) and offers approximately 3x the longitudinal stiffness of T6 aluminium, with typical longitudinal tensile modulus around 134–138 GPa [9]. The goal of this system being for the 3D printing parts to be the conduit of the forces of the arm for the carbon fiber tubes to absorb.

The reset of the arm attachment and design is ultimately to facilitate such a load path. The design heavily uses the 3D printings ability to create complex interlocking mechanisms, such as pins, pilot holes and fillets to create a firm attachment, even without the carbon fiber tubing. 4 6M bolts secure each end of the arm links in order to ensure no slippage from use over time, and to add another layer of protection. Given the sometimes fickle nature of 3D printing, I wanted to pursue a maximalist approach to my joint design to make sure no catastrophic failure happens during operation.

The length of the arm links were largely dictated by the 3D printer, which in this case is a Bambu X1C which has a 256mm^3 printing volume. With this in mind, I chose for the arm links between J2->J3 and J3->J4 to be 240mm in length. This still achieves a 1m arm length with the addition of the joints, which add to the total arm length.

 Insertion Hub of arm link for J3	 Arm Link End

Chapter Five: Control System

Control System

The CineWave control system is architected around a distributed intelligence model in which a host-level trajectory planner delegates real-time servo regulation to six independent motor controllers, each running closed-loop field-oriented control at 30 kHz. This separation of concerns reflects a deliberate design choice: the host computer is responsible for high-level decisions such as trajectory interpolation, user input processing, and kinematic solving, while the motor controllers autonomously manage current regulation, position tracking, and fault protection at timescales far faster than any software loop on the host could sustain. The result is a system that remains dynamically stiff and responsive even when the host is momentarily occupied with computationally expensive tasks such as inverse kinematics evaluation or communication with the digital twin pipeline.

Communication between the host and the servo controllers occurs over a CAN-FD (Controller Area Network with Flexible Data Rate) bus, a protocol originally developed for automotive powertrain systems and well-suited to the deterministic, low-latency requirements of multi-axis robotic control. CAN-FD extends the classical CAN 2.0 specification by permitting data payloads of up to 64 bytes at a switched bit rate of 5 Mbps during the data phase, while retaining the 1 Mbps arbitration phase that ensures collision-free bus access among multiple nodes. The CineWave system partitions its six axes across two independent CAN-FD channels, each driven by a dedicated MCP2518FD controller on the Waveshare 2-Channel Isolated CAN FD HAT. Channel one serves joints one through three—the base, shoulder, and elbow—while channel two serves joints four through six at the wrist. This partitioning halves the per-channel bus utilization and ensures that high-frequency wrist updates do not compete for bandwidth with the high-torque proximal joints, whose control frames carry larger payloads due to higher-resolution position and torque fields.

The host controller is a Raspberry Pi 5. The Pi communicates with the CAN FD HAT over two SPI buses, each mapped to one CAN channel via device tree overlays that instantiate the mcp251xfd kernel driver. Because the Linux kernel's default bit-timing calculator produces suboptimal segment parameters at 5 Mbps on 40 MHz clock CAN controllers, the CineWave system specifies manual bit timings (propagation, phase, and synchronization jump width segments) derived from the moteus reference documentation and validated empirically during commissioning. This detail, though small, proved critical: without corrected data-phase timings, the MCP2518FD transceiver was unable to synchronize to frames transmitted by the moteus controllers, manifesting as persistent bus-off errors that were initially difficult to diagnose.

Motor Controllers and Feedback

Each of the six CineWave axes is driven by a moteus r4.11 brushless servo controller paired with an mj5208 outrunner motor. The moteus r4.11 is a compact, open-source field-oriented control (FOC) driver capable of 100 A peak phase current. Position feedback is provided by an onboard magnetic encoder that senses a diametrically magnetized target affixed to the motor rotor. The encoder operates at the full 30 kHz control rate, providing sub-degree electrical angle resolution.

Power distribution follows a centralized topology. A MEAN WELL RSP-1500-24 switch-mode power supply converts mains AC to a 24 V DC bus at up to 62.5 A, feeding the mjbots power distribution board (r4.5b), which provides six individually switched and current-monitored output channels. Each channel supplies one moteus controller through an XT30 connector. The choice of 24 V as the bus voltage represents a compromise between the torque headroom available at higher voltages and the safety margin for operating the system in an educational and prototyping environment; the moteus r4.11 supports up to 44 V, leaving room for a future voltage increase should the CineWave's payload or speed requirements grow.

Torque feedback, reported by each moteus controller as part of its standard CAN-FD telemetry frame, which comes in handy for safety and homing of the joints. During the homing sequence, torque feedback becomes the primary sensing modality: the controller commands a slow constant-velocity motion toward a mechanical hard stop while monitoring the reported torque. When the joint makes contact with the stop, the motor velocity drops to near zero while the torque rises to the command limit, producing a distinctive stall signature that the homing algorithm detects and confirms over a configurable dwell time. This approach eliminates the need for dedicated limit switches or index sensors, reducing both wiring complexity and the number of failure points in the system.

Kinematics and Motion Planning

The CineWave arm is a serial manipulator whose kinematic structure closely resembles an anthropomorphic industrial robot with a spherical wrist. This decomposition into a positioning subchain and an orientation subchain is kinematically advantageous because it permits the inverse kinematics problem to be partitioned: given a desired camera pose in Cartesian space, the wrist center position can be computed by subtracting the wrist-to-camera offset from the desired position, the first three joint angles can then be solved geometrically to place the wrist center, and finally the last three joint angles can be extracted from the residual rotation matrix. This closed-form approach avoids the convergence issues and computational cost of iterative numerical solvers, making it feasible to evaluate at the 50 Hz control rate of the host loop.

The forward kinematics are expressed using the Denavit-Hartenberg (DH) convention, which assigns a local coordinate frame to each joint and expresses the transformation between consecutive frames as a product of four elementary transformations parameterized by the DH parameters: link length, link twist, link offset, and joint angle. For the CineWave, the DH table is straightforward: joints one through three have nonzero link lengths corresponding to the physical distances between joint axes, while joints four through six have zero link lengths and intersecting axes, consistent with the spherical wrist geometry. The joint offsets for the wrist joints account for the mechanical home position's alignment with the DH convention's zero configuration. The resulting homogeneous transformation from the base frame to the end-effector frame is the sequential product of six four-by-four matrices, each a function of a single joint variable, yielding a compact analytical expression for the camera's position and orientation as a function of the joint state vector.

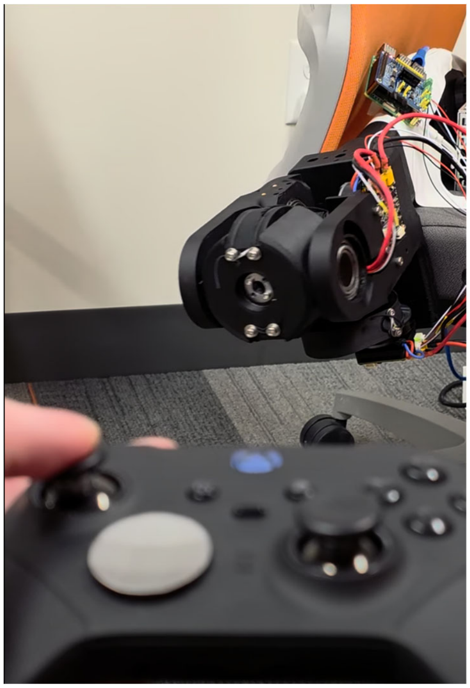
Motion planning for the CineWave operates at two levels. At the lower level, each moteus controller independently enforces acceleration and velocity limits on its joint trajectory, generating smooth trapezoidal or S-curve profiles between commanded waypoints. This local trajectory generation runs at the full servo rate and guarantees jerk-bounded motion even if the host issues step commands. For live operation via the Xbox controller, the host bypasses the keyframe interpolator and instead maps user inputs directly to joint velocity or position targets, relying on the per-axis trajectory limiters to ensure smoothness.

User Interface

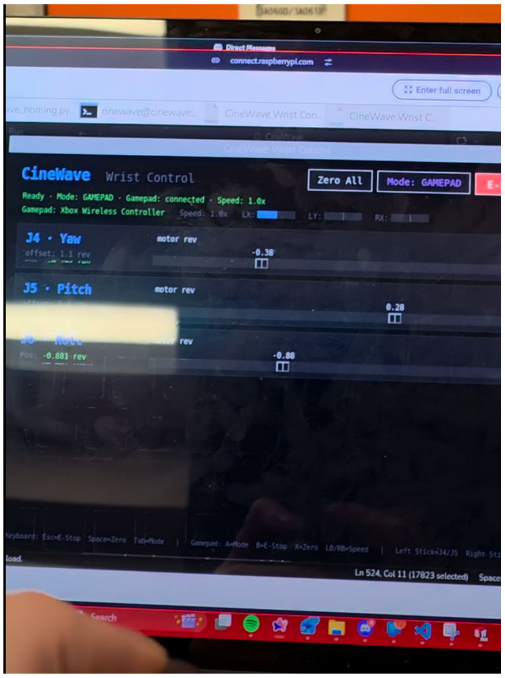
The CineWave currently supports two simultaneous input modalities: a graphical slider interface for precise numeric positioning and an Xbox Elite wireless controller for intuitive real-time puppeteering.

This dual-mode design reflects the observation that different phases of a cinematographic production demand fundamentally different interaction paradigms: shot programming and fine adjustment reward the precision and repeatability of direct numeric control, while live rehearsal and dynamic shot discovery benefit from the proprioceptive immediacy of a handheld controller whose physical affordances (analog stick deflection, trigger pressure, button state) map naturally to robotic motion parameters.

The Xbox controller communicates with the Raspberry Pi 5 over Bluetooth and is read through the Linux evdev input subsystem via the python-evdev library, which exposes raw axis and button events without requiring a graphical display server. A dedicated gamepad polling thread normalizes the raw 16-bit axis values to a signed unit range.



Movement control of the arm with a controller



```
Ready · Mode: GAMEPAD · Gamepad: connected · Speed: 1.0x
Gamepad: Xbox Wireless Controller   Speed: 1.0x   LX:   LY:   RX:
```

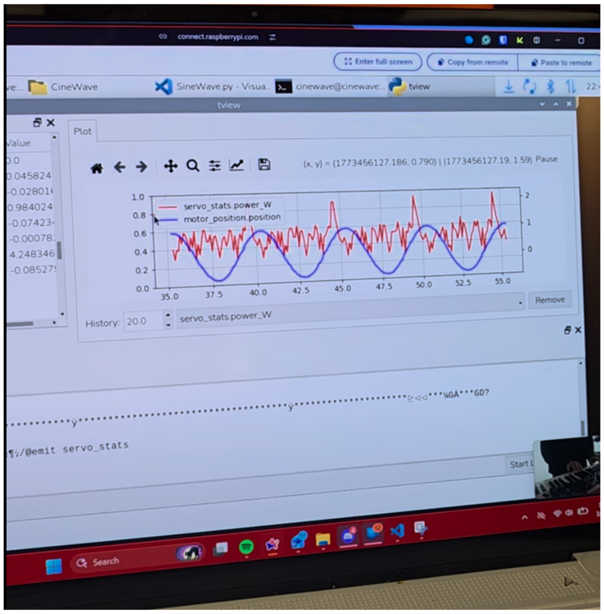
UI system to view position of each joint

The mapping between controller axes and robot joints follows a spatial metaphor intended to feel intuitive to camera operators familiar with two-axis pan/tilt heads: the left analog stick controls wrist yaw (horizontal axis) and wrist pitch (vertical axis), while the right analog stick's horizontal axis controls wrist roll. This assignment places the two most cinematographically expressive degrees of freedom (pan and tilt) on the dominant left stick, where the operator has the finest motor control. The stick-to-joint mapping uses a velocity-integration scheme rather than a direct position mapping: stick deflection sets a velocity proportional to the deflection magnitude, and the control loop integrates this velocity into a position target each cycle. The resulting behavior mimics a rate-controlled gimbal, where releasing the stick causes the camera to hold its current orientation rather than snapping back to center. A speed multiplier, adjustable in real time via the controller's bumper buttons, scales the maximum velocity from 0.1 to 3.0 times the default, allowing the operator to trade between fine adjustment precision and rapid

Real-time telemetry (position, velocity, and torque) is displayed alongside each slider at a 25 Hz refresh rate, and color coding provides at-a-glance status: green for normal operation, red for a faulted axis. An emergency stop button, mirrored by the controller's B button and the keyboard's Escape key, commands an immediate cessation of torque on all axes and latches until explicitly cleared, providing a consistent safety mechanism regardless of which input device the operator is holding at the moment of an emergency.

Chapter Six: Assembly and Integration

As of April 15, 2025, the arm is fully assembled and mostly functional, making it a major physical milestone for CineWave. The remaining work to be done is primarily in PID tuning and software design. A major software integration I aim to add is using Sonys Remote Capture SDK to the Raspberry Pi, allowing for a live, wireless stream from the camera mounted to the arm, with recording and modifications to things like ISO, aperture, and focus distance. This allows for a user to see what the camera is recording, but also opens up possibilities for automatic control using computer vision models running on the raspberry pi.

 A side profile of the robot arm mounted to a tripod	 A snapshot of my backlash test using the encoder on the motor controller.

The electrical backbone across all six axes is fully prepared. The MEAN WELL RSP-1500-24 power supply has been terminated and wired to the mjbots power distribution board (r4.5b), feeding 24V bus power to six moteus r4.11 controllers via XT30 connectors. CAN-FD signal wiring is soldered between all controllers and split across two bus segments.

On the software side, I've written over 1,200 lines of Python control code covering a complete homing sequence for all six joints and a real-time teleoperation interface using an Xbox controller. The Xbox controller mapping gives the operator direct joint-space velocity commands through the analog sticks and triggers: intuitive, two-handed control over six degrees of freedom. Communication with the moteus controllers runs through the Python moteus library's transport.cycle() method, batching position and velocity commands to all controllers in a single CAN-FD frame cycle for deterministic timing.

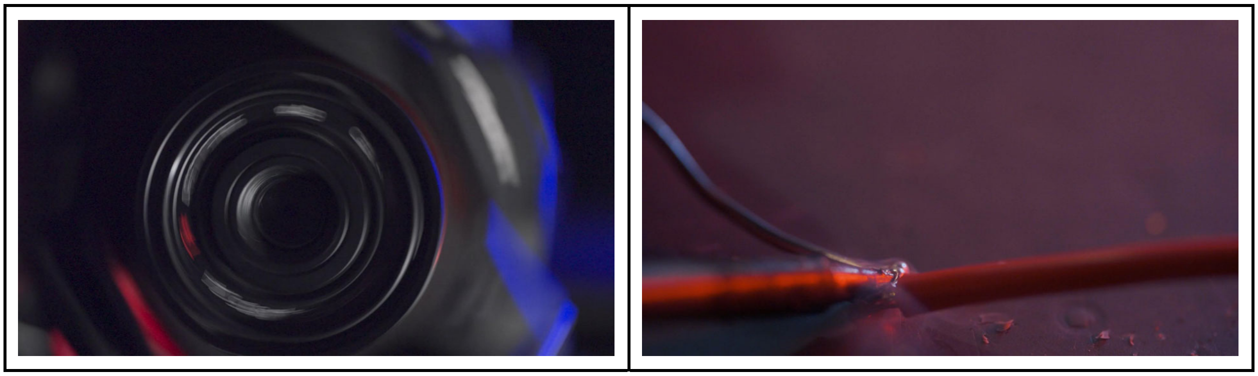
The primary bottleneck right now is print time. PPA-CF structural components (particularly the J2 and J3 housings) take upward of 24 hours per part at the 80 mm/s speeds and 50+% infill density I need for structural integrity. Print orientation has been selected for each component to align the stronger X-Y fiber direction (168 MPa tensile strength) with the primary load paths, since Z-direction strength drops to roughly 57 MPa due to the interlayer adhesion limitations inherent to short-fiber-reinforced FDM. In other words: the printer knows which way the forces go, even if it takes a full day to act on it.

What remains before the live capstone demonstration: completing the fabrication and assembly of J1–J3, where the HD-17 and HD-25 II harmonic drives serve as both speed reduction elements and primary structural bearings at each joint; building a digital twin of the arm in Unreal Engine 5.6 using the Interchange Framework GLB import pipeline, with Sequencer-based pre-visualization and a Live Link bridge from the Pi 5 for real-time pose mirroring; mounting a Sony α7 IV to the J6 Arca-Swiss interface and bolting the arm base to a pedestal tripod; and performing the live demonstration for capstone review.

Chapter Seven: Video & Unreal Engine Renders

As part of my TRF capstone class, I decided that, beyond simply building the arm, I would aim for a complementary video that goes over my philosophy on the project and an example of the arm in Unreal Engine. This culminated in in two videos: one 3-minute video of me working in the lab, done with a voiceover, and a 60-second Unreal Engine advertisement of the robot arm.

The IRL (In Real Life) video was directed and shot by me over the course of the assembly process, aiming to create a visually striking audio-visual journal of my journey. It features videos of my construction progress, dramatic recreations of the high and lows of my assembly, and examples of the robot arm in action.



The Unreal Engine sequence was started by exporting the joint assembly in SOLIDWORKS as GLB files. These were then positioned in Blender as a whole arm, with optimizations to the meshes and material properties for a more visually dynamic model that more closely resembles the real-life counterpart. I used the armature creation process in Blender to create the kinematics. Then I exported it as a GLB file into Unreal Engine.

The Unreal Engine sequence was a mix of surrealistic environments and realistic movement and was designed to maximize the visual aesthetic of the piece. I went deep into the sequencer to create smooth movement of each joint as the arm moves through the environment, showcasing the capabilities of the robotic arm. It is complemented by the classical song Winter Vivaldi.

Chapter Eight: Conclusion and Future Work

Summary of Contributions

This project's central contribution is a proof of concept: that cinema-grade robotic camera motion does not require cinema-grade budgets, and the future of media technology lies in indie productions. At roughly $2,500 in materials (about three percent of what an MRMC BOLT costs before you even talk about crew) CineWave achieves sub-pixel positioning accuracy at 1080p resolution and 5-meter subject distances, the threshold below which mechanical error becomes invisible to an audience.

The engineering strategy that makes this possible is the hybrid drive architecture. Rather than applying the same transmission technology uniformly across all six joints, whose which maximalist design philosophy has extremely diminished ROI, CineWave matches each joint to the cheapest mechanism that meets its specific precision and torque demands. Harmonic drives at J1 through J3, where sub-7-arcminute backlash tolerances and 60+ Nm torques leave no room for compromise. Antagonistic Dyneema cable capstans at J4 and J5, where the 30-arcminute budget opens the door to a lighter, cheaper solution that still eliminates backlash through constant tension rather than precision gearing. Direct drive at J6, where the torque requirement drops low enough that a speed reducer becomes unnecessary weight.

The structural design pushes a similar logic. PPA-CF 3D printing handles the complex geometries, cable channels, hollow shafts, and interlocking pin joints. These features would be prohibitively expensive to CNC from aluminum, while pultruded carbon fiber tubes running through the arm links carry the actual bending loads at roughly three times the longitudinal stiffness of T6 aluminum. This division of labor between additive and composite manufacturing is, as far as I can tell, uncommon in open-source robotics at this scale, and it is what allows a 3D printer with a 256mm build volume to produce structural components for a meter-long arm that doesn't flex under a 1.5 kg payload.

On the controls side, CineWave demonstrates that a $75 Raspberry Pi 5, six moteus r4.11 controllers, and a CAN-FD bus architecture borrowed from automotive powertrain systems can coordinate six-axis motion at rates sufficient for smooth cinematographic movement. The Xbox controller teleoperation interface makes controlling the robot more intuitive for newcomers. Camera operators think in terms of continuous motion, not discrete positions, and the controller's physical affordances map naturally to that mental model.

What CineWave contributes to the broader conversation is not a finished product. It is evidence that the gap between a $200,000 professional cinema robot and a $300 desktop arm is not an engineering inevitability but rather a design choice, and one that can be made differently when you start from the shot instead of the spec sheet. The impossible shot shouldn't be impossible because of your budget.

Potential Extensions (computer vision tracking, teach mode, etc.)

CineWave in its current form is a functional proof of concept, and certainly not a complete product or fully featured robotic arm. The path from here to something a film student can reliably pull out of a case and shoot with on a Tuesday involves targeted upgrades to the mechanical, electrical, and software systems; each informed by what building and testing the first prototype revealed.

The most immediate priority is structural. PPA-CF handled the complex geometries well, but the joint housings at J1 and J2 — the two joints that bear the full cantilevered load of the arm, payload, and their own dynamic forces — are operating closer to their material limits than I'm comfortable with. Future revisions will replace the most stressed housings with machined aluminum components, not wholesale, but at the specific interfaces where cyclic loading and bolt preload demand a material that doesn't creep under sustained stress the way any thermoplastic eventually will. The 3D-printed architecture stays everywhere it works. Where it doesn't, aluminum steps in. Same design philosophy as the hybrid drive strategy: match the material to the actual demand.

The motors at J1 and J2 are the other mechanical bottleneck. The mj5208 outrunners do the job at the current payload and speed targets, but they leave limited headroom for faster movements or heavier camera configurations, and their solid shaft design forces all cabling to route externally around the joint. Upgrading to higher-torque motors with central bore holes solves both problems at once: more torque margin for dynamic shots, and a coaxial cable path straight through the rotation axis, which eliminates the cable wrap and strain relief headaches that are currently the ugliest part of the wiring.

On the sensing side, the current magnetic encoders read position at the motor rotor, before the reduction stage. This means any backlash or compliance in the harmonic drive or capstan sits between the sensor and the actual joint position. The controller thinks the joint is somewhere it isn't, and the error is exactly the kind of thing that shows up as drift on screen. Adding encoders on the bearing side of each joint, after the transmission, closes that loop. The controller would know where the output actually is, not where the motor thinks it should be, and can correct for cable stretch, gear compliance, and thermal drift in real time.

The software roadmap splits into two directions. The first is object tracking: feeding the Sony camera's live video stream through a lightweight detection model running on the Pi 5 to keep a subject centered in frame as the arm moves. The inverse kinematics pipeline is already in place; the missing piece is a perception layer that converts pixel-space tracking data into Cartesian target updates at a rate fast enough to feel responsive. The second is building out the Unreal Engine 5.6 digital twin into a full previsualization and programming environment, where a cinematographer can choreograph complex multi-axis moves in virtual space, scrub through them, adjust timing, and then push the motion data to the physical arm for execution.

Bibliography

[1] M. B. Yilmaz, E. Lotman, A. Karjus, and P. Tikka, “An embodiment of the cinematographer: emotional and perceptual responses to different camera movement techniques,” *Front. Neurosci.*, vol. 17, p. 1160843, 2023, doi: 10.3389/fnins.2023.1160843.

[2] Mark Roberts Motion Control, “Bolt | High-Speed Cinema Robot.” [Online]. Available: https://www.mrmoco.com/motion-control/bolt/

[3] DJI, “DJI Ronin 2 Specifications.” [Online]. Available: https://www.dji.com/ronin-2/info

[4] “Thor – An Open Source 3D Printable 6DOF Robotic Arm.” Accessed: Nov. 30, 2025. [Online]. Available: http://thor.angel-lm.com/

[5] “Robotic Arm Classification: Degrees of Freedom (DoF) – The Smart Warehouse.” Accessed: Nov. 30, 2025. [Online]. Available: https://thesmartwarehouse.org/topics/robotic-arm-classification-degrees-of-freedom-dof/

[6] Harmonic Drive, “Product Types and Their Advantages.” [Online]. Available: https://www.harmonicdrive.net/technology/product-type-advantages

[7] Sony, “ILME-FX3 Specifications.” [Online]. Available: https://www.sony.co.uk/electronics/support/camcorders-and-video-cameras-interchangeable-lens-camcorders/ilme-fx3/specifications

[8] “Dynamica DM20.” [Online]. Available: https://dynamica-ropes.com/products/dm20/

[9] “Carbon Fiber Tubes Explained.” [Online]. Available: https://www.easycomposites.co.uk/learning/carbon-fibre-tubes-explained