Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Four arm-mounted piezoelectric harvesters on a quadrotor: a verified energy account and its consequences for energy-aware deep reinforcement learning navigation

Sayeed Omar

sayeedomar359@gmail.com

Tianjin Sino-German University of Applied Sciences

Ma Guoli

Tianjin Sino-German University of Applied Sciences

Research Article

Keywords: Piezoelectric energy harvesting, Quadrotor UAV, Finite-element analysis, Deep reinforcement learning, Energy-aware navigation, Reward shaping

Posted Date: November 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11416917/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Four arm-mounted piezoelectric harvesters on a quadrotor: a verified energy account and its consequences for energy-aware deep reinforcement learning navigation

Sayeed Omar[1,*] · Ma Guoli[1]

[1] *Department of Aviation and Aerospace Engineering, Tianjin Sino-German University of Applied Sciences, Tianjin 300350, China*

[*] Corresponding author: Sayeed Omar, sayeedomar359@gmail.com, ORCID 0009-0006-5708-7196

Abstract

Piezoelectric patches bonded to the four arms of a multirotor, combined with deep reinforcement learning (DRL) policies that favour high-harvest flight, have been proposed as a route to longer endurance. This paper tests that proposal with a single energy account in which structure, transducer, propulsion and policy share physical units. A finite-element (FE) model of a DJI F450 arm carrying its motor tip mass and a PZT-5A patch, verified against closed-form solutions and reproduced by an independent implementation, shows that the tip mass lowers the first bending frequency from 57.68 Hz to 18.43 Hz; neglecting it overestimates the mission-average harvest about 28-fold. With the tip mass, four root-mounted patches deliver 0.70 μW averaged over a representative mission (0.42 mJ in 10 min, 3.3×10^{-9} of propulsion energy). Even an arm resonance-tuned to the hover blade-passing frequency raises the four-arm hover output only to 21 μW (1.4×10^{-7} of hover power), and a ±5% change in take-off mass removes more than half of it. The FE harvest model and a rotor-speed-dependent propulsion model are then embedded in a mapless 3-D obstacle-avoidance task. Over five seeds and 200 held-out scenarios, SAC reaches the highest success (83.6 ± 3.6%, against 72.4% for PPO and 59.5% for DQN), whereas PPO uses 15.5% less energy per metre, so no algorithm dominates. An energy term in the reward lowers SAC energy per metre by 13.2% without loss of success; a harvest bonus does not increase harvesting and erodes most of that saving (+12.6%, Holm-adjusted p = 0.039). Endurance is gained by energy-aware flight, not by harvest-seeking policies.

Keywords Piezoelectric energy harvesting · Quadrotor UAV · Finite-element analysis · Deep reinforcement learning · Energy-aware navigation · Reward shaping

1 Introduction

Battery energy limits the endurance of electric multirotors to tens of minutes, and almost all of that energy is spent keeping the vehicle in the air: a 1.3 kg F450-class quadrotor draws on the order of 150 W in hover. Two largely separate research communities try to stretch this budget. The first harvests vibration energy from the airframe with piezoelectric transducers [1–4]; the second trains DRL policies that navigate cluttered spaces autonomously [10–14], occasionally with energy consumption in the objective. Recently the two have been combined, with arm-mounted harvesting feeding a DRL-controlled mission [9].

A quadrotor offers four nominally identical arms, each loaded by its own rotor, so a four-arm deployment is the natural way to scale such a harvester. Whether the coupling with navigation matters, however, depends on whether the harvested and propulsive energy flows are comparable, or whether the harvested power is at least large enough to change a decision the policy makes. Answering this needs one consistent energy account in which the structural model, the transducer model, the propulsion model and the navigation policy share the same physical units. Two modelling choices make that account easy to get wrong. A clamped bare-beam model of the arm omits the motor and propeller mass at the tip and therefore misplaces the arm modes relative to the rotor harmonics; and a harvest reward term that is rescaled to be "visible" to the policy can interact with the other reward terms in ways unrelated to harvesting.

This paper builds that account for a DJI F450 carrying a PZT-5A patch on each of its four arms and asks four questions:

- Q1. How much power can a surface-bonded patch on one arm deliver across the rotor-speed range once the motor/propeller tip mass is modelled, and how much do four arms deliver together?
- Q2. How does that power compare with propulsion power over a mission and with in-flight measurements reported for real F450 airframes?
- Q3. How far can the four-arm output be raised by wiring and by tuning the arm to resonance, and how robust is a tuned design?
- Q4. Does an energy-aware DRL navigation policy benefit from a reward term that encourages high-harvest flight, and what does energy awareness itself buy?

The contributions are:

- An open Euler–Bernoulli FE electromechanical model of the arm with the motor/propeller tip mass, verified against closed-form solutions and reproduced by a second, independently written implementation. Omitting the tip mass raises the first bending frequency from 18.43 Hz to 57.68 Hz and overestimates the mission-average harvest by a factor of ≈28.
- A four-arm energy budget across operating points, patch locations, load resistance, wiring and parameter uncertainty. Four root patches supply 3.3×10^{-9} of mission propulsion energy; even published in-flight measurements on a modified F450 [1] correspond to only 2.5×10^{-5}.
- An upper-bound analysis of resonance-tuned arms. Tuning either the first or the second arm mode to a hover harmonic raises the four-arm hover output to 21–91 μW (at most 6.2×10^{-7} of hover power); the gain is lost for take-off-mass changes of a few per cent, and the stiffer tuning costs about 10^5 times more propulsion power than it recovers.
- An energy-aware navigation benchmark in which propulsion power is computed from rotor speed and the four-arm FE harvest is evaluated at every step. SAC, PPO and DQN are compared over five seeds on an identical held-out test set, with a map-based A* planner as a privileged reference; all analytic gradients are verified by finite differences.
- A controlled test of harvest-bonus reward shaping, showing how it changes rotor speed, harvested power and propulsion energy, and why it fails.

All code, raw results and the 25 unit tests used in this study are released with the paper (see Data availability).

2 Related work

Vibration harvesting on UAVs. Anton and Inman [3] demonstrated piezoelectric and photovoltaic harvesting on a small fixed-wing UAV, and Erturk et al. [4] modelled an L-shaped beam–mass harvester for UAV applications, building on distributed-parameter cantilever models [5] and earlier generator designs for wireless electronics [6]. On multirotors, Perez et al. [1] redesigned the four arms of a DJI F450 to carry commercial patches and buzzers. In hover they reported up to 5.35 mW for the four arms (extrapolated from measurements on one arm), enough for low-power structural-health sensors but, in their words, equivalent to extending a 10-min flight by only about 0.2 s. Gonçalves et al. [2] bench-tested three series-connected PVDF films on a motor–arm assembly and obtained about 4 μW through an

LTC3588-1 conditioner [8]. Magnetically coupled and nonlinear harvesters that broaden the bandwidth are reviewed in [7].

DRL navigation. Deep Q-networks [15], proximal policy optimisation [16, 17] and soft actor-critic [18, 19] are the usual algorithms for mapless UAV obstacle avoidance [25], whether the input is vision [10, 11], fused LiDAR and depth [12] or low-dimensional range data [13]; DRL policies have also reached champion level in drone racing [14]. Reported comparisons are often based on a single seed or on differing evaluation sets, which makes differences hard to interpret [21, 22]. Bahi and Ourici [9] combined DRL with piezoelectric harvesting towards self-sustaining drone operations. The present work is complementary: it asks, with an explicitly verified four-arm energy account, how large the harvested term can be relative to propulsion and what a harvest-aware reward actually changes.

3 Four-arm harvester model

3.1 Structure and excitation

Each arm is modelled as a clamped–free Euler–Bernoulli beam of length L, width b and thickness h that represents the glass-fibre-reinforced nylon arm by an equivalent solid rectangular section (Table 1). It carries a lumped tip mass $m_{\text{tip}} = 73$ g for the motor, propeller and fasteners. The beam is discretised with 80 Hermite-cubic elements and consistent mass matrices; the harmonic response is obtained by superposing the first eight mass-normalised modes with modal damping $\zeta = 0.02$.

Rotor thrust follows $T = C_T \rho n^2 D^4$ with n in rev/s [24]. The vertical dynamic load at the arm tip is represented by its two dominant harmonics. The first, at the shaft frequency (1P), comes from blade-to-blade thrust asymmetry and track error; the second, at the blade-passing frequency (2P for a two-bladed propeller), comes from thrust ripple:

$$F(t) = \delta_{1P} T \sin(2\pi n t) + \delta_{2P} T \sin(4\pi n t), \ \delta_{1P} = 0.01, \ \delta_{2P} = 0.02 \quad (1)$$

A rotating mass imbalance acts in the rotor plane and excites lateral rather than vertical bending of the arm. The resulting strain is antisymmetric across the width of a patch centred on the top surface, so it produces no net charge and is not included. The ripple fractions δ are assumptions; because power scales with δ^2, they are varied in the sensitivity study (Sect. 4.5).

3.2 Electromechanical coupling

A PZT-5A patch of length L_p, width b_p and thickness t_p is bonded to the upper surface of each arm with its centre at x_c. In the weak-coupling limit ($t_p/h = 0.033$, so the patch changes neither the arm stiffness nor its modes appreciably) the short-circuit charge amplitude is proportional to the patch-averaged surface strain, which integrates exactly to a difference of FE nodal slopes θ:

$$Q = d_{31} Y_p b_p z_p [\theta(x_2) - \theta(x_1)], \;\; z_p = \frac{h + t_p}{2} \tag{2}$$

The patch is a current source $j\omega Q$ in parallel with its clamped capacitance $C_p = \varepsilon^S_{33} L_p b_p / t_p = 49.8$ nF, feeding a resistive load R:

$$V = \frac{j\omega Q R}{1 + j\omega R C_p}, \;\; P = \frac{\omega^2 |Q|^2 R}{2[1 + (\omega R C_p)^2]}, \;\; R_{opt} = \frac{1}{\omega C_p}, \;\; P_{max} = \frac{\omega |Q|^2}{4 C_p} \tag{3}$$

The 1P and 2P components are at different frequencies, so their time-averaged powers add. Voltages, currents and powers are therefore mutually consistent by construction.

Table 1 Model parameters. The arm is an equivalent rectangular section; rotor coefficients are representative of a 9.4-in two-bladed propeller.

Symbol	Arm and patch	Value	Symbol	Rotor and vehicle	Value
L	free length	245 mm	D	propeller diameter	0.239 m
$b \times h$	section	16 × 6 mm	C_T, C_P	thrust, power coeff.	0.11, 0.045
E, ρ	GF-nylon	18.5 GPa, 1450 kg/m^3	η	motor + ESC efficiency	0.75
m_{tip}	motor + prop + mount	73 g	m	take-off mass	1.30 kg
ζ	modal damping	0.02	P_{av}	avionics	6 W
$L_p \times b_p \times t_p$	PZT-5A patch	50 × 15 × 0.2 mm	n_{max}	maximum speed	8,200 rpm
d_{31}	piezo. coefficient	−171 pm/V	δ_{1P}	1P thrust asymmetry	0.01
Y_p	patch modulus	61.0 GPa	δ_{2P}	2P thrust ripple	0.02
$\varepsilon^T_{33}/\varepsilon_0$	relative permittivity	1700			

3.3 Four-arm configuration

One patch is bonded at the same location on each of the four arms. The arms are treated as identical and uncoupled, i.e. frame-level modes and hub compliance are neglected (Sect. 7.3). With rotor speeds n_i, the total harvested power is

$$P_h = \sum_{i=1}^{4} P_{patch}(n_i) = 4 P_{patch}(n) \tag{4}$$

for equal rotor speeds n, when each patch drives its own optimal load or conditioner. If the four patches are instead wired together and excited in phase, the total power at the respective optimal

load is the same, and wiring changes only the voltage and the optimal load: a series string has four times the open-circuit voltage, capacitance $C_p/4$ and optimal load $4R_{opt}$; a parallel connection has capacitance $4C_p$ and optimal load $R_{opt}/4$. Because the 2P ripples of the four rotors are not phase-locked, these coherent-sum values are upper bounds.

3.4 Propulsion power

Electrical power per rotor and total propulsion power are

$$P_{rot} = \frac{C_P \rho n^3 D^5}{\eta}, \quad P_{prop} = 4P_{rot} + P_{av} \tag{5}$$

with the parameters in Table 1. The hover condition $4T = mg$ gives a hover speed of 5,109 rpm and a hover power of 147.6 W for the 1.30 kg vehicle. This is consistent with typical F450 figures, although it is not calibrated against a specific motor–propeller test.

3.5 Verification

Without the tip mass, the FE model reproduces the closed-form cantilever frequencies 57.68 Hz and 361.45 Hz, with a relative error of 8.5×10^{-6} for f_1. These bare-arm values are what a model without the motor predicts. The static tip deflection equals $FL^3/3EI$ to machine precision, and the fundamental frequency with the tip mass (18.43 Hz) agrees with a Rayleigh estimate that adds 33/140 of the beam mass to the tip. Frequencies converge to five significant figures at 20 elements. The electrical identities in Eq. (3) are checked numerically: the load sweep recovers $R_{opt} = 1/(\omega C_p)$ within 1%, $V \to V_{oc}$ as $R \to \infty$, and $P(R_{opt}) = \omega|Q|^2/4C_p$. These checks are part of the unit-test suite (Appendix B). In addition, a second implementation written independently from the equations above reproduces every frequency, strain-derived voltage and power in Tables 2–4 to the precision reported.

4 Harvester results

4.1 Modes and operating points

With the tip mass, the first three bending modes lie at 18.43, 259.7 and 827.0 Hz (Fig. 1a). Both the 1P (85–137 Hz) and the 2P (170–273 Hz) excitation over the operating range lie above f_1, so the arm responds mass-controlled, with small strain, except where the 2P harmonic approaches f_2. That happens at 7,792 rpm, where the root patch peaks at 17.0 μW (Fig. 1c). Table 2 lists strains, open-circuit voltages and powers at four operating points. At hover the patch-averaged strain amplitude is only 0.25 με and the power 0.029 μW.

Table 2 Root-patch response at four operating points (design load 13.9 kΩ). Values separated by "/" are the 1P and 2P harmonics.

Operating point	rpm	f (Hz)	Strain (με)	V_{oc} amplitude (mV)	R_{opt} (kΩ)	P (μW)
Hover	5,109	85.2 / 170.3	0.25 / 0.26	40 / 42	22.6	0.029
Cruise	6,000	100.0 / 200.0	0.27 / 0.42	43 / 67	17.7	0.079
Climb	7,000	116.7 / 233.3	0.31 / 1.04	48 / 163	13.9	0.503
Maximum	8,200	136.7 / 273.3	0.36 / 2.17	57 / 342	11.8	2.486

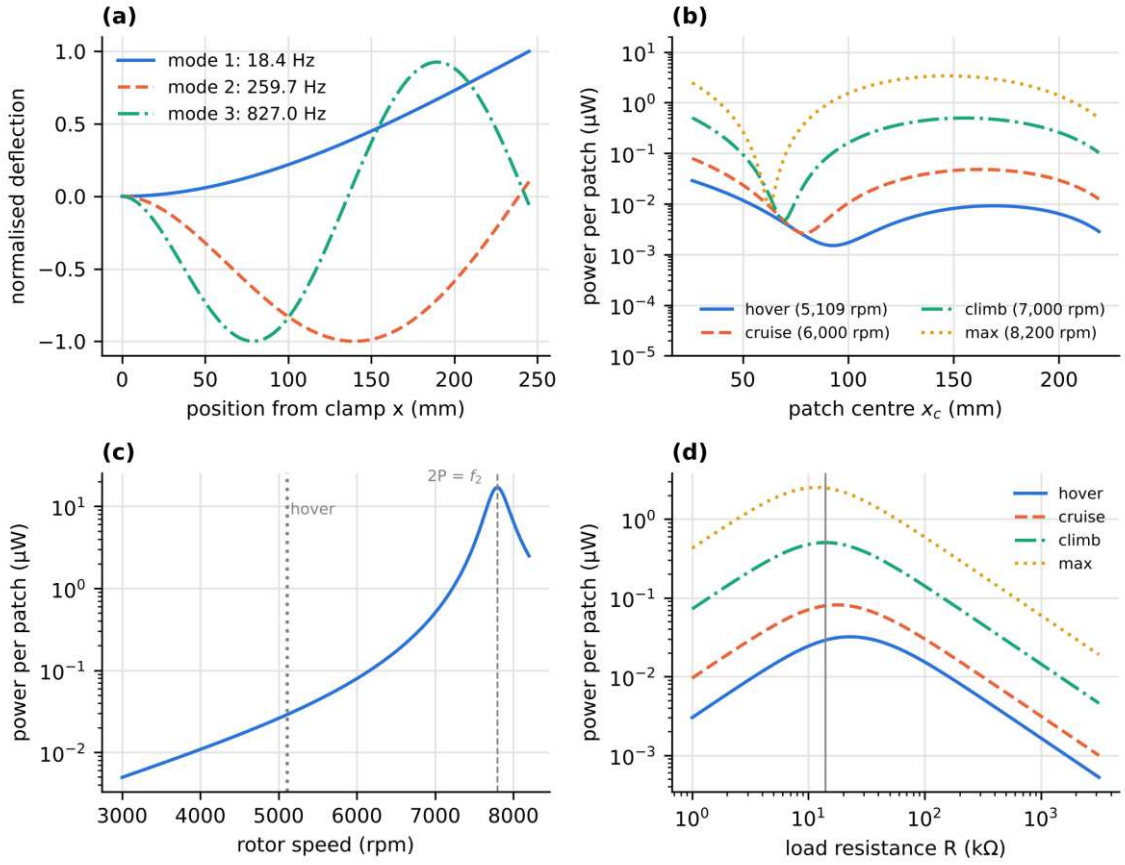


Fig. 1 FE results for one F450 arm with motor tip mass. (a) First three bending modes. (b) Power per patch versus patch-centre position at the four operating points (design load R = 13.9 kΩ). (c) Root-patch power versus rotor speed; the peak is the 2P harmonic crossing the second mode. (d) Power versus load resistance at the root patch; the vertical line marks the design load.

4.2 Patch location and load

Figure 1b sweeps the patch centre along the arm. The best location depends on which harmonic dominates. At hover and cruise the root (patch starting 1 mm from the clamp) harvests most because the quasi-static first-mode curvature is largest there. Near maximum speed a mid-span patch collects slightly more because it sits on the curvature antinode of the second mode. Averaged over a representative mission profile (55% hover, 30% cruise at 6,000 rpm, 12% climb at 7,000 rpm and 3% at maximum speed), the root location is best (Table 3), delivering 5.4 times more than a patch next to the motor mount. The optimal load falls from 22.6 kΩ at

hover to 11.8 kΩ at maximum speed, following $1/(\omega C_p)$ for the dominant harmonic (Fig. 1d). A fixed 13.9 kΩ load maximises mission-average power.

Table 3 Power per patch (μW) at three patch locations (design load 13.9 kΩ).

Location	x_c (mm)	Hover	Cruise	Climb	Maximum	Mission average
Root (P-R)	26.0	0.029	0.079	0.503	2.486	0.175
Mid-span (P-M)	122.5	0.004	0.028	0.354	2.801	0.137
Motor mount (P-T)	219.0	0.003	0.012	0.103	0.495	0.032

4.3 Four-arm energy budget

Table 4 scales the root-patch results to four arms and compares them with the propulsion power of Eq. (5) at the same rotor speed. Over the mission profile four patches deliver 0.70 μW, or 0.42 mJ in 10 min, while the rotors consume 128 kJ (average 214 W); the harvested fraction is 3.3×10^{-9}. Even at the 2P–f_2 coincidence (7,792 rpm) the four-arm output of 68 μW is 1.3×10^{-7} of propulsion power. For comparison, the 5.35 mW reported for hover by Perez et al. [1] on arms enlarged to carry several transducers corresponds to 2.5×10^{-5} of the mission-average propulsion power. Either figure is orders of magnitude too small to affect endurance (Fig. 2b).

Table 4 Four-arm harvest and propulsion power (root patches, 13.9 kΩ load per patch).

Condition	rpm	Per patch (μW)	Four arms (μW)	P_{prop} (W)	Harvest / P_{prop}
Hover	5,109	0.029	0.116	147.6	7.8×10^{-10}
Cruise	6,000	0.079	0.317	235.3	1.3×10^{-9}
Climb	7,000	0.503	2.01	370.1	5.4×10^{-9}
Maximum	8,200	2.486	9.94	591.2	1.7×10^{-8}
2P = f_2 peak	7,792	17.0	68.0	508.1	1.3×10^{-7}
Mission average	–	0.175	0.70	213.9	3.3×10^{-9}

4.4 Wiring and power conditioning

Conditioning is a further constraint. The LTC3588-1 used in several UAV studies [2] operates from a 2.7–20 V input and draws 450 nA in under-voltage lockout [8], i.e. about 1.2 μW at its minimum operating input. This exceeds the mission-average output of all four patches (0.70 μW). The open-circuit amplitudes of a single patch (40/42 mV for the 1P/2P components at hover, 57/342 mV at maximum speed) are far below the start-up voltage. Connecting the four patches in series raises the voltage at most fourfold (Sect. 3.3): the coherent peak sum is ≈0.33 V at hover and ≈1.6 V at maximum speed, still below 2.7 V. A practical four-arm supply would therefore need a low-voltage cold-start converter or operation deliberately close to a structural resonance (Sect. 4.6).

4.5 Sensitivity

Table 5 varies the least certain inputs one at a time. The tip mass matters most: removing it (the bare-beam case) raises the mission-average power to 4.95 μW, ≈28 times the baseline, because the modes move towards the excitation frequencies. Thinner arms behave similarly. Damping and the 1P asymmetry have a minor effect, and power scales roughly with δ_{2P}^2 because the 2P harmonic dominates above cruise speed. Across all cases the harvested power stays within 0.04–5 μW per patch, which does not change the conclusion of Sect. 4.3.

Table 5 One-at-a-time sensitivity of the modal frequencies and the mission-average root-patch power (load re-optimised for each case).

Parameter	Value	f_1 (Hz)	f_2 (Hz)	Mission-average P (μW)
Baseline (Table 1)	–	18.43	259.7	0.175
Tip mass m_{tip} (g)	0	57.68	361.5	4.950
Tip mass m_{tip} (g)	50	21.77	262.6	0.398
Tip mass m_{tip} (g)	100	15.96	258.0	0.094
Arm thickness h (mm)	4	10.20	171.7	2.572
Arm thickness h (mm)	8	27.91	349.1	0.044
Arm thickness h (mm)	10	38.40	439.7	0.047
Damping ratio ζ	0.01	18.43	259.7	0.184
Damping ratio ζ	0.04	18.43	259.7	0.148
2P thrust ripple δ_{2P}	0.01	18.43	259.7	0.051
2P thrust ripple δ_{2P}	0.04	18.43	259.7	0.669
1P thrust asymmetry δ_{1P}	0.005	18.43	259.7	0.167
1P thrust asymmetry δ_{1P}	0.02	18.43	259.7	0.206

4.6 Upper bound: resonance-tuned arms

The baseline output is small because both excitation harmonics lie away from the arm modes. The most favourable case for four-arm harvesting is therefore an arm designed so that a mode coincides with a hover harmonic. Within the equivalent-section model, two such tunings are possible by changing the arm thickness alone.

Second mode at the hover 2P frequency. Thinning the arm to h = 3.97 mm places f_2 at 170.35 Hz. The root patch then delivers 5.30 μW at its optimal load of 18.8 kΩ, and four arms deliver 21.2 μW, i.e. 1.4×10^{-7} of hover power (12.7 mJ in a 10-min hover; Fig. 2b). The thinner arm has 71% less bending stiffness, and its static tip deflection under hover thrust grows from 2.9 mm to 10.1 mm. Even if four such patches were wired in series and excited in phase, the coherent peak voltage would be about 4 × (37 + 630) mV ≈ 2.7 V, only at the lower edge of the LTC3588-1 input range.

First mode at the hover 1P frequency. Thickening the arm to h = 17.7 mm places f_1 at 85.2 Hz and gives 22.9 µW per patch (optimal load 37.5 kΩ), or 91 µW for four arms (6.2×10^{-7} of hover power). The four thicker arms, however, add about 0.27 kg, and because hover power scales as $m^{3/2}$ in this rotor model, hover power rises by roughly 30% (≈45 W), about 5×10^{5} times more than is harvested.

Resonant gain is also narrow. With ζ = 0.02 the half-power bandwidth is about ±2% in frequency, which corresponds to ±2% in rotor speed or ±4% in thrust. Figure 2c shows the four-arm hover output of the f_2-tuned arm versus take-off mass: a ±2% change retains 78–82% of the tuned output, ±5% retains 36–42% and ±10% only 12–17%. Payload changes, battery swaps and attitude control routinely shift rotor speed by this much, so a tuned arm would rarely operate at its design point.

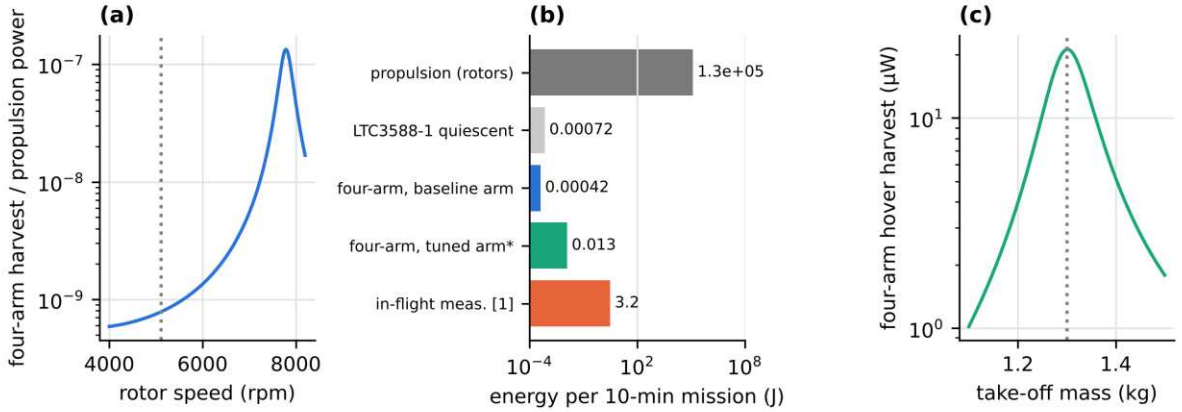


Fig. 2 Four-arm energy account. (a) Four-arm harvest (root patches, 13.9 kΩ) as a fraction of propulsion power versus rotor speed; dotted line: hover. (b) Energy over a 10-min mission (log scale): propulsion, LTC3588-1 under-voltage-lockout quiescent draw, four baseline arms, four f_2-tuned arms in hover (*), and the in-flight measurement of [1]. (c) Hover output of four f_2-tuned arms (h = 3.97 mm, 18.8 kΩ per patch) versus take-off mass; dotted line: design mass.

5 Energy-aware navigation benchmark

5.1 Task and dynamics

The vehicle must fly from a start point near one wall of a 10 × 10 × 5.5 m volume to a goal near the opposite wall (straight-line distance 7–9 m) through eight spherical obstacles of radius 0.5–0.8 m, half of which are placed near the direct line. Each scenario is generated from an integer seed. The vehicle is modelled as a point mass integrated at Δt = 0.05 s. The action is a commanded acceleration $a \in [-2, 2]^3$ m/s^2; the velocity obeys $\mathrm{d}v/\mathrm{d}t = a - 0.3v + w$ with a speed cap of 4 m/s, where w is an optional gust acceleration (Sect. 5.5). An episode ends with success when the goal is within 0.5 m, with a crash when the body (radius 0.25 m) touches an obstacle

or leaves the volume, or with a time-out after 400 steps (20 s). The policy observes 17 values: the unit vector and clipped distance to the goal, the velocity, and ten range readings (eight horizontal rays at 45° spacing plus up and down, maximum range 3 m). It has no map.

5.2 Energy and harvest accounting

At every step the thrust magnitude $T = m\|a + g\mathbf{e}_z\|$ is shared equally by the four rotors. Inverting $T/4 = C_T\rho n^2 D^4$ gives the rotor speed, from which the propulsion power P_{prop} (Eq. 5) and the four-arm harvest power $P_h = 4P_{patch}(n)$ (Eq. 4) follow. P_{patch} is interpolated from the FE root-patch curve (Fig. 1c) at the 13.9 kΩ design load. Both energies are integrated over the episode. This simplification ignores attitude dynamics and differential thrust but keeps the key coupling: aggressive manoeuvres raise rotor speed, which raises both propulsion power ($\propto n^3$) and harvest power.

5.3 Reward configurations

The reward at step t is

$$r_t = 5(d_{t-1} - d_t) - 0.01 - \lambda_E \frac{P_{prop}}{P_{prop}^{hover}} + \lambda_H \frac{P_h}{P_h^{hover}} + 10 \cdot 1_{goal} - 10 \cdot 1_{crash} \qquad (6)$$

where d_t is the distance to the goal. Three configurations are compared: *standard* ($\lambda_E = \lambda_H = 0$), *energy-aware* ($\lambda_E = 0.1$, $\lambda_H = 0$) and *energy-aware + harvest bonus* ($\lambda_E = \lambda_H = 0.1$). If harvested energy were credited in the same physical units as propulsion energy, the harvest term would be at most 2.3×10^{-9} of the energy term; numerically it would be indistinguishable from the energy-aware reward, so that variant needs no experiment. The harvest bonus is therefore normalised by its hover value, as in Eq. (6), which gives it the same weight as the energy penalty. This is deliberately the most favourable test of the idea that a policy should seek high-harvest flight.

5.4 Algorithms

SAC with automatic temperature tuning [18, 19], PPO with a clipped surrogate and generalised advantage estimation [16, 17], and DQN with a target network and Huber loss [15] were implemented in NumPy from the original formulations. The hyperparameters, listed in full in Appendix A, are close to common library defaults [20]; they were chosen once and not tuned per algorithm or reward configuration. DQN uses the 27 actions $\{-2, 0, 2\}^3$ m/s^2. Every analytic gradient, including the SAC reparameterised actor gradient through the tanh squashing and the

PPO clipped-surrogate gradient with respect to the log standard deviation, is checked against central finite differences (Appendix B).

5.5 Protocol and statistics

Each configuration is trained with five seeds (0–4) for 200,000 environment steps. Every 10,000 steps the deterministic policy is evaluated on 30 validation scenarios; these curves are reported but not used for model selection. The final policy of each run is evaluated deterministically on the same 200 held-out test scenarios, which are also used for the reference planner, so the per-seed success rate has a resolution of 0.5 percentage points (pp). The energy metric is propulsion energy per metre of straight-line goal distance (J/m), averaged over successful episodes. Seeds are the unit of replication: results are reported as mean ± standard deviation over seeds, pairwise differences are tested with Welch's t-test with Holm correction, effect sizes are Hedges' g, and three-group comparisons also use one-way ANOVA. With five seeds per group these tests have limited power, so effect sizes and per-seed values are reported as well [21, 22].

Robustness is assessed without retraining: the five energy-aware SAC policies are evaluated on the first 100 test scenarios with Gaussian noise on the normalised range readings (σ = 0.05, 0.10), with Ornstein–Uhlenbeck gust accelerations (time constant 2 s, stationary standard deviation σ = 0.5, 1.0 m/s^2), and with both combined.

Reference planner. A 3-D A* search [23] on a 0.2 m grid (obstacles inflated by 0.5 m, 26-connected, Euclidean heuristic) is followed by a PD tracker that steers towards a look-ahead point 1.5 m along the path, $a = 2.0(p_{\text{look}} - p) - 1.5v$, clipped to the same actuator limits. The gains were chosen on the validation scenarios as the lowest-energy setting with 100% validation success. The planner knows the exact obstacle map and its own position, information the DRL policies do not have, so it serves as a privileged reference rather than a competing method.

6 Navigation results

6.1 Algorithm comparison

Figure 3a shows the learning curves and Table 6 the test results with the energy-aware reward. SAC learns fastest and reaches the highest test success, 83.6 ± 3.6%, against 72.4 ± 6.2% for PPO and 59.5 ± 13.4% for DQN (ANOVA F = 9.42, p = 0.0035). After Holm correction SAC differs from both PPO and DQN (p = 0.035 for each, Hedges' $g \geq 2$), while the PPO–DQN

difference is not significant (p = 0.10; Table 8). Nearly all failures are collisions rather than time-outs.

Energy gives a different ranking. PPO policies fly faster (2.46 m/s against 2.09 m/s for SAC) and therefore spend less time aloft, which lowers their energy per metre by 15.5% relative to SAC; they collide more often, though. Figure 4a shows that the per-seed results form a trade-off rather than a single winner: no learned policy is best on both success and energy. The map-based A*+PD reference succeeds in every test scenario, as expected from its privileged information. It flies more cautiously (1.57 m/s), and energy-aware SAC uses 10.1% less energy per metre than it does.

Table 6 Test results on 200 held-out scenarios, mean ± s.d. over five seeds (energy, time and speed are averaged over successful episodes). The A*+PD planner is deterministic and has access to the exact map.

Method	Success (%)	Crash (%)	Time-out (%)	Energy (J/m)	Flight time (s)	Speed (m/s)
SAC (mapless, energy-aware)	83.6 ± 3.6	16.2 ± 3.7	0.2 ± 0.3	85.7 ± 2.8	4.9 ± 0.2	2.09 ± 0.10
PPO (mapless, energy-aware)	72.4 ± 6.2	27.5 ± 6.3	0.1 ± 0.2	72.4 ± 4.9	4.0 ± 0.2	2.46 ± 0.10
DQN (mapless, energy-aware)	59.5 ± 13.4	33.8 ± 5.4	6.7 ± 9.7	112.9 ± 9.7	6.2 ± 0.5	1.76 ± 0.07
A*+PD (exact map, reference)	100.0	0.0	0.0	95.3	5.5	1.57

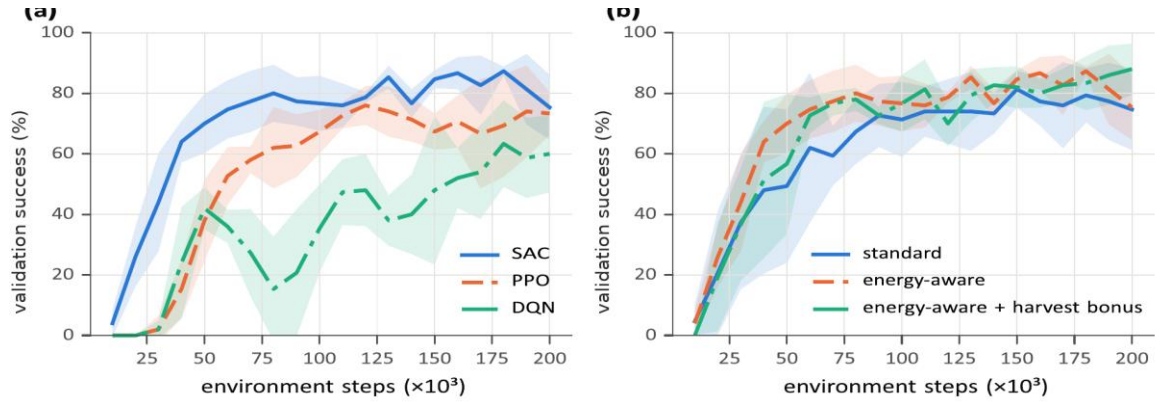


Fig. 3 Validation success during training (mean ± s.d. over five seeds; 30 validation scenarios every 10^4 steps). (a) SAC, PPO and DQN with the energy-aware reward. (b) SAC with the three reward configurations.

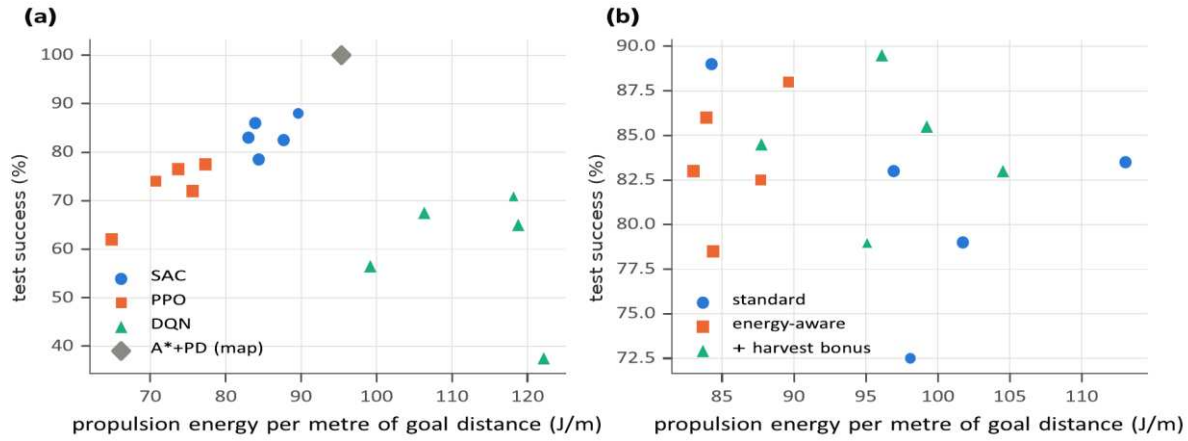


Fig. 4 Per-seed test success versus propulsion energy per metre of goal distance. (a) Algorithms with the energy-aware reward, and the A*+PD reference. (b) SAC with the three reward configurations.

6.2 Effect of the energy and harvest terms

Table 7 compares the three SAC reward configurations. None of them changes success (ANOVA $p = 0.60$). The energy term lowers propulsion energy per metre from 98.8 ± 10.3 to 85.7 ± 2.8 J/m, a 13.2% reduction with a large effect size ($g = 1.56$). With five seeds this is not significant after Holm correction ($p = 0.090$), but the three-group ANOVA on energy is ($F = 4.82$, $p = 0.029$), and energy-aware seeds are also much less variable. The saving comes mainly from time: hover power dominates, so the energy penalty acts as an additional time cost, and the energy-aware policies fly faster (2.09 against 1.91 m/s) and more directly (Fig. 5).

Adding the harvest bonus does not make the policy seek high-harvest flight. The mean rotor speed and the harvested power are statistically unchanged (Table 8), because within the actuator limits the rotors stay near hover speed, where the harvest curve is flat and small (Fig. 1c). What the bonus does change is the time cost. At hover the normalised bonus (+0.1 per step) exactly cancels the normalised energy penalty (−0.1 per step), so the net incentive to finish quickly falls back to that of the standard reward. The harvest-bonus policies accordingly fly as slowly as the standard ones (1.88 m/s) and use 12.6% more energy per metre than energy-aware SAC ($p = 0.039$, $g = 2.05$), losing most of the saving. In physical terms the question is moot: an energy-aware SAC episode harvests 0.61 µJ with all four arms while spending 738 J on propulsion, a ratio of 8.2×10^{-10}.

Table 7 SAC with three reward configurations (five seeds, 200 test scenarios). Harvest power is the four-arm total averaged over the flight time of successful episodes.

Reward	Success (%)	Energy (J/m)	Speed (m/s)	Mean rotor speed (rpm)	Four-arm harvest (µW)
Standard (no energy term)	81.4 ± 6.1	98.8 ± 10.3	1.91 ± 0.13	5,132 ± 36	0.123 ± 0.004
Energy-aware	83.6 ± 3.6	85.7 ± 2.8	2.09 ± 0.10	5,133 ± 53	0.124 ± 0.006

Reward	Success (%)	Energy (J/m)	Speed (m/s)	Mean rotor speed (rpm)	Four-arm harvest (µW)
Energy-aware + harvest bonus	84.3 ± 3.8	96.5 ± 6.1	1.88 ± 0.08	5,096 ± 29	0.120 ± 0.003

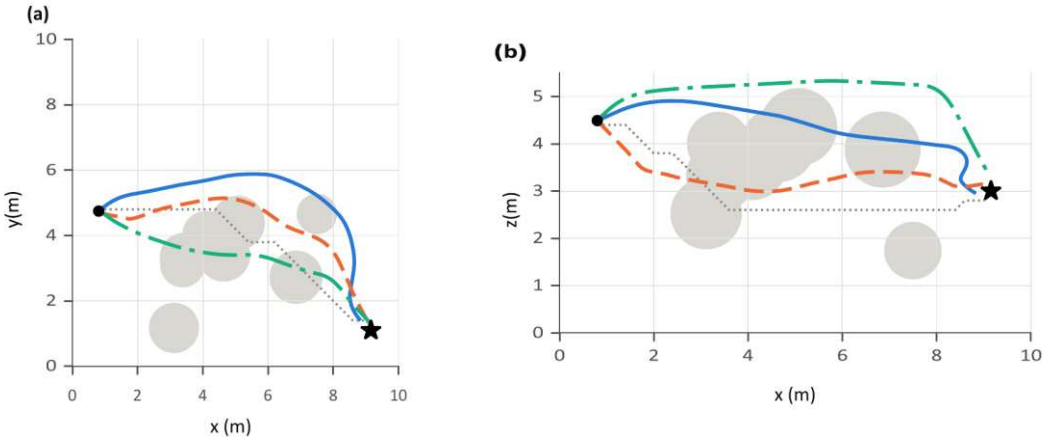


Fig. 5 Example trajectories (seed-0 policies) in the first test scenario in which all three SAC variants succeed, in (a) top and (b) side projection. Solid blue: SAC standard (1168 J, 8.0 s); dashed orange: SAC energy-aware (995 J, 6.7 s); dash-dotted green: SAC + harvest bonus (902 J, 6.2 s); dotted grey: A* path on the exact map. Values are propulsion energy and flight time. Obstacles are drawn as their projections, so apparent overlaps with paths are projection effects.

Table 8 Statistical comparisons over seeds (n = 5 per group): Welch's t-test, Holm-adjusted p within each metric, Hedges' g. Δ is the first group minus the second.

Metric	Comparison	Δ	t / F	p	p_{Holm}	g
Algorithms (energy-aware reward)						
Success (pp)	SAC vs PPO	+11.20	3.49	0.012	0.035	1.99
Success (pp)	SAC vs DQN	+24.10	3.88	0.014	0.035	2.22
Success (pp)	PPO vs DQN	+12.90	1.95	0.102	0.102	1.12
Success (pp)	ANOVA (3 groups)	–	F = 9.42	0.003	–	–
Energy (J/m)	SAC vs PPO	+13.28	5.27	0.002	0.003	3.01
Energy (J/m)	SAC vs DQN	−27.20	−6.00	0.002	0.003	−3.43
Energy (J/m)	PPO vs DQN	−40.47	−8.30	<0.001	<0.001	−4.74
Energy (J/m)	ANOVA (3 groups)	–	F = 50.42	<0.001	–	–
Reward configurations (SAC)						
Success (pp)	Standard vs Energy-aware	−2.20	−0.69	0.513	1.000	−0.40
Success (pp)	Standard vs Harvest-bonus	−2.90	−0.90	0.400	1.000	−0.51
Success (pp)	Energy-aware vs Harvest-bonus	−0.70	−0.30	0.774	1.000	−0.17
Success (pp)	ANOVA (3 groups)	–	F = 0.53	0.603	–	–
Energy (J/m)	Standard vs Energy-aware	+13.09	2.74	0.045	0.090	1.56
Energy (J/m)	Standard vs Harvest-bonus	+2.28	0.42	0.685	0.685	0.24
Energy (J/m)	Energy-aware vs Harvest-bonus	−10.82	−3.58	0.013	0.039	−2.05
Energy (J/m)	ANOVA (3 groups)	–	F = 4.82	0.029	–	–
Rotor speed (rpm)	Standard vs Energy-aware	−0.70	−0.02	0.981	0.981	−0.01
Rotor speed (rpm)	Standard vs Harvest-bonus	+35.94	1.73	0.124	0.373	0.99
Rotor speed (rpm)	Energy-aware vs Harvest-bonus	+36.64	1.35	0.224	0.447	0.77
Rotor speed (rpm)	ANOVA (3 groups)	–	F = 1.32	0.303	–	–
Harvest (µW)	Standard vs Energy-aware	−0.0018	−0.57	0.584	0.584	−0.33

Metric	Comparison	Δ	t / F	p	p_{Holm}	g
Harvest (µW)	Standard vs Harvest-bonus	+0.0029	1.37	0.209	0.490	0.78
Harvest (µW)	Energy-aware vs Harvest-bonus	+0.0047	1.60	0.163	0.490	0.91
Harvest (µW)	ANOVA (3 groups)	–	$F = 1.47$	0.268	–	–

6.3 Robustness

Without retraining, energy-aware SAC is insensitive to noise on the range readings (82.0% success at $\sigma = 0.10$ against 82.8% nominal on the same 100 scenarios; Table 9). Gusts are more damaging: success falls to 74.8% at σ = 0.5 m/s^2 and 54.8% at σ = 1.0 m/s^2. The A*+PD reference, whose tracker was tuned without wind, degrades by a similar amount (76.0% and 48.0%). Disturbance rejection, not perception noise, is therefore the main weakness of both approaches in this benchmark.

Table 9 Success (%) under perturbations on the first 100 test scenarios (SAC: mean ± s.d. over five energy-aware policies, no retraining).

Condition	SAC (mapless)	A*+PD (exact map)
Nominal	82.8 ± 5.1	100.0
Ray noise $\sigma = 0.05$	81.0 ± 6.9	100.0
Ray noise $\sigma = 0.10$	82.0 ± 6.5	100.0
Wind σ = 0.5 m/s^2	74.8 ± 2.0	76.0
Wind σ = 1.0 m/s^2	54.8 ± 4.9	48.0
Noise 0.05 + wind 0.5	76.2 ± 3.7	75.0

7 Discussion

7.1 What the numbers imply

The harvested power of an arm-bonded patch is set by the arm's strain, and the strain is small because the heavy motor at the tip places the first mode well below both excitation harmonics. The bare-beam simplification removes exactly this effect and inflates the prediction. Adding arms scales the output linearly, but four times a negligible quantity remains negligible: the four-arm mission average is 3.3×10^{-9} of propulsion energy, and even the most favourable published measurement, obtained with arms redesigned to carry several transducers [1], is more than four orders of magnitude below propulsion power. Resonance tuning does not change this picture. It raises the four-arm output by two to three orders of magnitude, but only within a few per cent of the design take-off mass, and the stiffer of the two tunings costs far more propulsion power than it recovers. Arm-mounted harvesting is therefore a sensor-supply technology, possibly useful for duty-cycled structural-health monitoring of the arms themselves [1], and not an endurance technology.

For navigation this has a direct consequence. Credited in physical units, a harvest term cannot affect a learned policy. Scaled up until it can, it distorts the other reward terms: here it neutralised the energy penalty instead of steering the vehicle towards high-harvest flight. What does reduce energy is making the policy value time and thrust, which the energy-aware reward achieves at no cost in success.

7.2 Practical recommendations

(1) Include the motor/propeller mass in structural models of multirotor arms; a clamped bare beam overestimates harvest by more than an order of magnitude.

(2) Report open-circuit voltage, load, current and power together, so that they can be checked against $P = V^2/2R$, and compare the harvest with the quiescent draw and start-up voltage of the intended conditioner.

(3) Compare four-arm harvested energy with propulsion energy over the same mission, and account for the mass and stiffness changes any harvester-driven redesign of the arms implies.

(4) Keep energy-related reward terms in physical units, or check explicitly how their normalisation interacts with the other terms.

7.3 Limitations

The arm is represented by an equivalent solid rectangular section; real F450 arms are hollow and stiffer, which would raise the modal frequencies (compare the h = 8–10 mm rows of Table 5). The dynamic thrust fractions δ_{1P} and δ_{2P} are assumed, and frame-level modes, hub compliance, inter-arm coupling and in-plane loads are not modelled. Together these probably explain why the in-flight measurement of [1] exceeds the model prediction. The four arms are treated as identical and uncoupled, and coherent series addition is an upper bound. The harvester predictions have not been validated experimentally in this work. A bench test that clamps an arm with its motor and propeller, sweeps 3,000–8,200 rpm and records surface strain and patch voltage across a matched load would test Table 2 and Fig. 1c directly; a four-arm frame test would add the coupling effects omitted here. The navigation study uses a point-mass model with a hover-based power estimate that ignores attitude dynamics, forward-flight induced-velocity effects and battery voltage sag, so absolute energies should be read as relative indicators. Hyperparameters were not tuned per algorithm, which may disadvantage PPO and DQN in particular. The algorithms are custom NumPy implementations, verified by gradient

checks but not benchmarked against reference libraries. Finally, five seeds per configuration give limited statistical power, so effect sizes and per-seed data are reported alongside the *p*-values.

8 Conclusions

A verified finite-element model of a DJI F450 arm with its motor tip mass predicts that a root-mounted PZT-5A patch harvests 0.029 μW at hover and 0.175 μW averaged over a mission. Four arms together recover 0.42 mJ in 10 min, 3.3×10^{-9} of propulsion energy; omitting the tip mass overestimates this by a factor of about 28. Tuning the arms to resonance raises the four-arm hover output to at most 21–91 μW, a gain that is lost for take-off-mass changes of a few per cent and that, for the stiffer design, is outweighed by the added arm mass. In a mapless obstacle-avoidance benchmark with physically computed propulsion and four-arm harvest power, SAC achieved the highest success (83.6%), PPO the lowest energy per metre, and no algorithm dominated both. An energy term in the reward reduced SAC propulsion energy by 13.2% without loss of success, whereas a harvest bonus did not increase harvesting and cancelled most of that saving. The practical conclusion is that endurance is gained by flying efficiently, not by seeking harvest. Future work will validate the structural model on the bench and on a four-arm frame in flight, use a full rigid-body quadrotor model with a calibrated motor–propeller map, and assess four-arm harvesting as a power source for duty-cycled onboard sensors.

Declarations

Funding No funding was received for conducting this study.

Competing interests The authors have no competing interests to declare that are relevant to the content of this article.

Ethics approval and consent Not applicable. This study involves no human participants or animals.

Author contributions Sayeed Omar: conceptualisation, methodology, software, formal analysis, investigation, visualisation, writing – original draft. Ma Guoli: supervision, writing – review and editing. Both authors read and approved the final manuscript.

Relation to other work Earlier versions of parts of this work were posted as preprints by the authors (engrXiv, https://engrxiv.org/preprint/view/6841; Preprints.org 202605.1916,

https://www.preprints.org/manuscript/202605.1916). Those versions used a clamped bare-beam arm model without the motor tip mass; their power estimates are superseded by the present manuscript.

Data and code availability The FE model, navigation environment, learning algorithms, unit tests, raw per-episode results (JSON) and figure scripts are available at Zenodo (https://doi.org/10.5281/zenodo.23446357) and are provided to reviewers as supplementary material. All results are reproduced by python3 fea_run.py, train.py (one call per run) and analyze.py; the only dependencies are NumPy, SciPy and Matplotlib.

Use of AI tools An AI assistant (Claude, Anthropic) was used under the authors' direction to help implement and verify the simulation code, including an independent re-implementation of the FE model, to run the simulations, and to draft and revise the manuscript text. The authors reviewed all code, results and text and take full responsibility for the content of this article.

References

[1] Perez M, Billon K, Gerges T, Capsal JF, Cabrera M, Chesné S, Jean-Mistral C (2022) Vibration energy harvesting on a drone quadcopter based on piezoelectric structures. Mech Ind 23:20. https://doi.org/10.1051/meca/2022021

[2] Gonçalves LS, Pereira RML, Tyszler RS, Morais MCAM, Barbosa CRH (2025) Application of PVDF transducers for piezoelectric energy harvesting in unmanned aerial vehicles. Energies 18:4759. https://doi.org/10.3390/en18174759

[3] Anton SR, Inman DJ (2008) Vibration energy harvesting for unmanned aerial vehicles. In: Proc SPIE 6928, Active and Passive Smart Structures and Integrated Systems 2008, 692824. https://doi.org/10.1117/12.776099

[4] Erturk A, Renno JM, Inman DJ (2009) Modeling of piezoelectric energy harvesting from an L-shaped beam-mass structure with an application to UAVs. J Intell Mater Syst Struct 20:529–544. https://doi.org/10.1177/1045389X08098096

[5] Erturk A, Inman DJ (2008) A distributed parameter electromechanical model for cantilevered piezoelectric energy harvesters. J Vib Acoust 130:041002. https://doi.org/10.1115/1.2890402

[6] Roundy S, Wright PK (2004) A piezoelectric vibration based generator for wireless electronics. Smart Mater Struct 13:1131–1142. https://doi.org/10.1088/0964-1726/13/5/018

[7] Jiang J, Liu S, Feng L, Zhao D (2021) A review of piezoelectric vibration energy harvesting with magnetic coupling based on different structural characteristics. Micromachines 12:436. https://doi.org/10.3390/mi12040436

[8] Analog Devices (2023) LTC3588-1 Piezoelectric energy harvesting power supply, data sheet. https://www.analog.com/en/products/ltc3588-1.html

[9] Bahi A, Ourici A (2026) Self-sustaining drone operations through deep reinforcement learning and piezoelectric energy harvesting. Int J Intell Robot Appl. https://doi.org/10.1007/s41315-025-00490-y

[10] Kalidas AP, Joshua CJ, Md AQ, Basheer S, Mohan S, Sakri S (2023) Deep reinforcement learning for vision-based navigation of UAVs in avoiding stationary and mobile obstacles. Drones 7:245. https://doi.org/10.3390/drones7040245

[11] Zhang Y, Yan C, Xiao J, Feroskhan M (2025) NPE-DRL: Enhancing perception constrained obstacle avoidance with nonexpert policy guided reinforcement learning. IEEE Trans Artif Intell 6:184–198

[12] Lei B, Hu W, Ren Z, Ji S (2025) DRL-based UAV autonomous navigation and obstacle avoidance with LiDAR and depth camera fusion. Aerospace 12:848. https://doi.org/10.3390/aerospace12090848

[13] Yin Y, Wang Z, Zheng L, Su Q, Guo Y (2024) Autonomous UAV navigation with adaptive control based on deep reinforcement learning. Electronics 13:2432. https://doi.org/10.3390/electronics13132432

[14] Kaufmann E, Bauersfeld L, Loquercio A, Müller M, Koltun V, Scaramuzza D (2023) Champion-level drone racing using deep reinforcement learning. Nature 620:982–987. https://doi.org/10.1038/s41586-023-06419-4

[15] Mnih V, Kavukcuoglu K, Silver D, et al (2015) Human-level control through deep reinforcement learning. Nature 518:529–533. https://doi.org/10.1038/nature14236

[16] Schulman J, Wolski F, Dhariwal P, Radford A, Klimov O (2017) Proximal policy optimization algorithms. arXiv:1707.06347

[17] Schulman J, Moritz P, Levine S, Jordan M, Abbeel P (2016) High-dimensional continuous control using generalized advantage estimation. In: Proc ICLR 2016. arXiv:1506.02438

[18] Haarnoja T, Zhou A, Abbeel P, Levine S (2018) Soft actor-critic: off-policy maximum entropy deep reinforcement learning with a stochastic actor. In: Proc ICML 2018, PMLR 80:1861–1870

[19] Haarnoja T, Zhou A, Hartikainen K, et al (2018) Soft actor-critic algorithms and applications. arXiv:1812.05905

[20] Raffin A, Hill A, Gleave A, Kanervisto A, Ernestus M, Dormann N (2021) Stable-Baselines3: reliable reinforcement learning implementations. J Mach Learn Res 22(268):1–8

[21] Henderson P, Islam R, Bachman P, Pineau J, Precup D, Meger D (2018) Deep reinforcement learning that matters. In: Proc AAAI 2018, pp 3207–3214

[22] Agarwal R, Schwarzer M, Castro PS, Courville A, Bellemare MG (2021) Deep reinforcement learning at the edge of the statistical precipice. In: Adv Neural Inf Process Syst 34, pp 29304–29320

[23] Hart PE, Nilsson NJ, Raphael B (1968) A formal basis for the heuristic determination of minimum cost paths. IEEE Trans Syst Sci Cybern 4:100–107. https://doi.org/10.1109/TSSC.1968.300136

[24] Leishman JG (2006) Principles of Helicopter Aerodynamics, 2nd edn. Cambridge University Press, Cambridge

[25] Sutton RS, Barto AG (2018) Reinforcement Learning: An Introduction, 2nd edn. MIT Press, Cambridge, MA

Appendix A Hyperparameters

Table A1 Learning hyperparameters (identical for all seeds and reward configurations).

	SAC	**PPO**	**DQN**
Network (hidden, activation)	2 × 128, ReLU	2 × 64, ReLU (separate π, V)	2 × 128, ReLU
Optimiser, learning rate	Adam, 3×10^{-4}	Adam, 3×10^{-4}	Adam, 3×10^{-4}
Discount γ	0.99	0.99	0.99
Batch / minibatch	128	64 (rollout 2048, 10 epochs)	64
Replay buffer	3×10^{5}	–	10^{5}
Warm-up (uniform random)	5,000 steps	–	5,000 steps
Target network	Polyak $\tau = 0.005$	–	hard copy every 1,000 steps
Algorithm-specific	auto-α (init. 0.1), target entropy −3; tanh-squashed Gaussian, log $\sigma \in [-5, 2]$	clip 0.2, GAE $\lambda = 0.95$, value coef. 0.5; state-independent log σ (init. 0)	train every 4 steps, Huber loss; ε: 1 → 0.05 over 20% of steps
Gradient-norm clipping	none	0.5	10
Training steps	200,000 per run; 5 seeds (0–4)	200,000 per run; 5 seeds (0–4)	200,000 per run; 5 seeds (0–4)

Appendix B Unit tests

All tests run with python3 test_unit.py; 25 of 25 pass. Gradient checks compare analytic gradients with central finite differences ($h = 10^{-6}$) on randomly chosen parameters and require a relative error below 10^{-4} (actor, policy) or 10^{-5} (critics, value, MLP).

Table B1 Unit-test suite.

#	**Test**	**Result**
1	FE f_1 (no tip mass) vs closed form < 0.01%	Pass
2	FE f_2 (no tip mass) vs closed form < 0.01%	Pass
3	FE static tip deflection = $FL^3/3EI$	Pass
4	FE f_1 with tip mass vs Rayleigh estimate < 1%	Pass
5	Electrical: numerical optimum load = $1/(\omega C_p)$	Pass
6	Electrical: $V(R \to \infty) = V_{oc}$	Pass
7	Electrical: $P(R_{opt}) = \omega\lvert Q\rvert^2/(4C_p)$	Pass
8	Rotor model: hover rpm reproduces hover thrust	Pass
9	Env: zero command → hover rpm	Pass
10	Env: ray–sphere range (+x) = 2.0 m	Pass
11	Env: ray to ceiling = min(3, 5.5 − 2)	Pass
12	Env: ray to floor = 2 − 0.3	Pass
13	Env: collision detected inside R + body radius	Pass
14	Env: success inside goal tolerance	Pass
15	Env: hover step energy = $P_{hover}\Delta t$	Pass
16	Env: observation dimension = 17	Pass
17	MLP parameter gradients (finite difference)	Pass

#	Test	Result
18	MLP input gradients (finite difference)	Pass
19	SAC critic 1 gradients (finite difference)	Pass
20	SAC critic 2 gradients (finite difference)	Pass
21	SAC actor reparameterised gradients incl. tanh correction (finite difference)	Pass
22	SAC log-prob = log $N(u) - \Sigma \log(1 - \tanh^2 u)$	Pass
23	PPO clipped-surrogate gradients incl. log-std (finite difference)	Pass
24	PPO value gradients (finite difference)	Pass
25	DQN Huber TD gradients (finite difference)	Pass