

Article

Two-Dimensional Thompson Sampling for Joint Beam and Power Control for Uplink Maritime Communications

Kyeong Jea Lee ¹, Joo-Hyun Jo ², Sungyoon Cho ³, Ki-Won Kwon ³ and DongKu Kim ^{1,*}¹ School of Electrical and Electronic Engineering, Yonsei University, Seoul 03722, Republic of Korea; kjlee92@yonsei.ac.kr² 6G Connected Mobility Standard TP, LG Electronics, Seoul 08826, Republic of Korea; joohyun.jo@lge.com³ Smart Network Research Center, Korea Electronics Technology Institute (KETI), Seoul 02456, Republic of Korea; sycho@keti.re.kr (S.C.); kwonkw@keti.re.kr (K.-W.K.)

* Correspondence: dkkim@yonsei.ac.kr

Abstract

In a cellular maritime communication system, ocean buoys are essential to enable environmental monitoring, offshore platform management, and disaster response. Therefore, energy-efficient transmission from the buoys is a key requirement to prolong their operational time. A fixed uplink beamforming can be considered to save energy by leveraging its beam gain while managing the target link reliability. However, the dynamic condition of ocean waves causes buoys' random orientation, leading to frequent misalignment of their predefined beam direction aimed at the base station, which degrades both the link reliability and energy efficiency. To address this challenge, we propose a wave-adaptive beamforming framework to satisfy data-rate demands within limited power budgets. This strategy targets scenarios where sea state information is unavailable, such as in network-assisted systems. We propose a Two-Dimensional Thompson Sampling (2DTS) scheme that jointly selects beamwidth and transmit power to satisfy the target-rate constraint with minimal power consumption and thus achieve maximal energy efficiency. This adaptive learning approach effectively balances exploration and exploitation, enabling efficient operation in uncertain and changing sea conditions. In simulation, under a moderate sea state, 2DTS achieves an energy efficiency of 1.26×10^4 bps/Hz/J at round 600, which is 73.7% of the ideal (1.71×10^4), and yield gains of 96.9% and 447.8% over exploration-based TS and conventional TS, respectively. Under a harsh sea state, 2DTS attains 3.09×10^4 bps/Hz/J (85.6% of the ideal 3.61×10^4), outperforming the exploration-based and conventional TS by 83.9% and 113.1%, respectively. The simulation results demonstrate that the strategy enhances energy efficiency, confirming its practicality for maritime communication systems constrained by limited power budgets.

Keywords: maritime communication; 2DTS; beam management; energy efficiency

Academic Editors: Jiaguo Liu and Jun Ye

Received: 30 September 2025

Revised: 19 October 2025

Accepted: 21 October 2025

Published: 23 October 2025

Citation: Lee, K.J.; Jo, J.-H.; Cho, S.; Kwon, K.-W.; Kim, D. Two-Dimensional Thompson Sampling for Joint Beam and Power Control for Uplink Maritime Communications. *J. Mar. Sci. Eng.* **2025**, *13*, 2034. <https://doi.org/10.3390/jmse13112034>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The growing demand for advanced maritime applications, such as surveillance, environmental monitoring, offshore platform management, and disaster response services, requires buoy-based ocean communication networks [1,2]. In particular, lighthouses and maritime buoys now play a critical role, going beyond simple signaling to transmitting real-time environmental and situational data including wave height, surrounding visibility, and environmental video feeds. These high data-rate requirements necessitate more effi-

cient wireless transmission strategies than the low-rate, omnidirectional links traditionally used for basic telemetry (e.g., wave height, air temperature).

Unlike terrestrial networks, maritime communication systems must deal with a highly dynamic ocean environment. Waves, wind and currents induce buoy motions that tilt, heave and rotate, increasing the probability of outage and energy consumption [3]. Although these effects are governed by the sea state, real-time sea state information is often unavailable. As a result, static power allocation schemes often result in inefficiencies in spectral and energy resources [4]. To address this, recent studies have explored adaptive techniques such as beamforming and multi-hop relaying to strengthen maritime links [5,6]. While directional beamforming can focus energy toward the intended receiver [7], most existing methods presume prior knowledge of sea conditions at the buoys, limiting their utility in the use cases when the sea conditions or channel state information are not available [8,9]. Buoys generally possess only partial and delayed knowledge of the prevailing sea state for uplink beamforming, whereas the base station (BS) can infer sea-state dynamics more reliably from satellite-based, wide-area forecasts. Moreover, it is noted that the instantaneous beamforming that adapts to the changing orientation of the buoy requires channel state information feedback from the base station, where cellular maritime communication such as Long-Term Evolution for Maritime (LTE-M) lacks a Multiple-Input Multiple-Output (MIMO) framework.

To address this challenge with respect to the uncertainty of the sea state and its related feedback channel unavailability in the current marine cellular system, energy-efficient uplink transmission becomes a crucial design goal, since maritime buoys operate under limited battery capacity and must maintain reliable connectivity under harsh sea-state variations. Recent studies have emphasized the importance of optimizing power efficiency and robustness in underwater and backscatter communication scenarios, such as in the Anti-Jamming Colonel Blotto Game for Underwater Acoustic Backscatter Communication [10]. Motivated by these challenges, we aim to minimize power consumption while satisfying the target rate constraint in uncertain sea-state environments. Therefore, we propose an adaptive beam and power control method based on Thompson Sampling (TS)-based reinforcement learning [11] without knowledge on the sea state directly from the base station. Because the base station cannot have knowledge on the sea state directly, we exploit the feedback information to implicitly infer the current sea state. To optimize beamwidth and power allocation under uncertain conditions, we employ the Thompson Sampling (TS) framework, a Bayesian algorithm widely used in multi-armed bandit (MAB) problems [12–14]. TS is known to efficiently balance exploration and exploitation by sampling from the posterior belief distributions of each action. In particular, the posterior is updated based on binary feedback success or failure, while allowing the algorithm to gradually converge to the most promising actions. TS has been successfully applied in areas such as online advertising, recommendation systems, and clinical trials. In communication systems, it has been used for channel selection, interference management, and power control, including underwater acoustic communications and dynamic spectrum access [14,15].

Unfortunately, most of the above-mentioned applications focus on maximizing performance metrics like throughput or reliability, with limited attention paid to energy efficiency [16]. Unlike these existing applications, in our scenario, beamforming performance is highly sensitive to sea-state conditions, where dynamic wave motion causes variations in signal alignment and channel quality. Alongside strict power constraints at buoys, this creates a multi-dimensional optimization problem involving both beamwidth and transmit power, which makes standard TS not suitable for our problem. This limitation arises because conventional Thompson Sampling (TS) is designed to maximize the expected

reward of a single action rather than to handle constrained or multi-dimensional objectives. If the reward is defined as the achieved rate, TS tends to converge toward high-power actions, since higher power monotonically increases the rate. Even when the reward is represented as a binary indicator of meeting the target rate, the algorithm still favors the high power that maximizes success probability, without accounting for power minimization. By contrast, our problem follows an objective: minimize transmit power subject to satisfying a predefined target rate, which cannot be directly encoded in the standard TS formulation. Moreover, unlike traditional TS that optimizes a single action, our setting requires joint exploration and exploitation over two coupled decision variables, beamwidth and transmit power, which form a two-dimensional action space. To address this challenge, we develop a Two-Dimensional Thompson Sampling (2DTS) approach that maintains beta posteriors for each beamwidth–power pair and applies a dominance-aware selection rule. Among all feasible pairs whose sampled success probabilities exceed the target reliability, 2DTS selects the configuration with the lowest power, thereby ensuring energy-efficient yet reliable transmission under uncertain sea-state conditions. The goal of 2DTS is to jointly select the optimal beamwidth and the minimum transmission power required to meet a predefined target data rate. This is achieved by updating the shape parameters α and β of the posterior beta distribution for each beamwidth and power pair selection actions, using binary feedback from the base station indicating whether the target rate was successfully achieved. This approach enables adaptive learning that accounts for underlying sea-state variations while minimizing power consumption. By integrating 2DTS into our framework, we provide an efficient solution that balances exploration and exploitation, ensuring reliable and energy-efficient communication in dynamic and uncertain maritime environments.

The paper is organized as follows. Section 2 models wave-induced fluctuations in the uplink transmit beam gain by incorporating buoy motion and sea-surface dynamics. Section 3 presents the Two-Dimensional Thompson Sampling (2DTS) algorithm, which jointly selects beamwidth and transmit power to meet the target-rate constraint with minimal power, thereby maximizing energy efficiency. Section 4 reports simulation results that validate the modeling and quantify the gains of the proposed approach. Finally, Section 5 concludes the paper.

2. System Model

As illustrated in Figure 1, we consider an uplink maritime communication scenario where a buoy equipped with N transmit antennas communicates with a single-antenna base station (BS). The buoy's location and orientation are assumed to be known, so the buoy employs predefined fixed transmit beamforming aiming at the base station without dynamic adjustments. The buoy transmits the scalar information symbol $x \sim \mathcal{CN}(0, P)$ to the BS through a linear beamforming vector $\mathbf{f} \in \mathbb{C}^{N \times 1}$. Accordingly, the received signal, denoted as y at the BS, is represented as

$$y = \mathbf{h}\mathbf{f}x + n, \quad (1)$$

where $\mathbf{h} \in \mathbb{C}^{1 \times N}$ is the channel between the buoy and BS, and $n \in \mathbb{C}$ follows an additive white Gaussian noise (AWGN) distribution [17]. For simplicity, $\mathbf{f}$ is chosen from a predefined vertical beam vector set $\mathcal{F} = \{\mathbf{f}_1, \dots, \mathbf{f}_5\}$, where they are the arrays steering to BS with different vertical beamwidths. It is assumed that the uplink beam vectors are designed to be horizontally isotropic and vertically directional because the buoy can frequently be rotating due to the ocean current. $\mathbf{f}_1$ indicates the beamforming vector with beamwidth $\theta_1 = \frac{\pi}{4}$, and subsequent beamwidths are progressively halved, given by

$$\theta_s = \frac{\pi}{2^{s+1}}, \quad s = 1, 2, \dots, 5, \quad (2)$$

corresponding to $\frac{\pi}{4}$, $\frac{\pi}{8}$, $\frac{\pi}{16}$, $\frac{\pi}{32}$, and $\frac{\pi}{64}$ [18]. For modeling the channel $\mathbf{h}$, following [19], we assume

$$\mathbf{h} = \sqrt{\frac{K}{K+1}} \mathbf{h}_{\text{LoS}} + \sqrt{\frac{1}{K+1}} \mathbf{h}_{\text{NLoS}}, \quad (3)$$

where $\mathbf{h}_{\text{LoS}}$ and $\mathbf{h}_{\text{NLoS}}$ represent the Line-of-Sight (LoS) and the Non Line-of-Sight (NLoS) components of the channel, respectively. For the NLoS component, we assume that each element in $\mathbf{h}_{\text{NLoS}}$ is drawn from $\mathcal{CN}(0, 1)$. This part captures the scattered multipath components. The Rician factor K represents the ratio of the power in the LoS component to the power in the NLoS component. The LoS component, $\mathbf{h}_{\text{LoS}}$, reflects the deterministic nature of the direct transmission path, which is particularly dominant in the maritime environment due to the lack of obstacles between the BS and the buoy [20]. Although the buoy is placed at known locations, the sea surface moves dynamically and causes buoy orientation changes, resulting in misalignment of the direction of the transmit array of the buoy and degradation of the gain at BS [21].

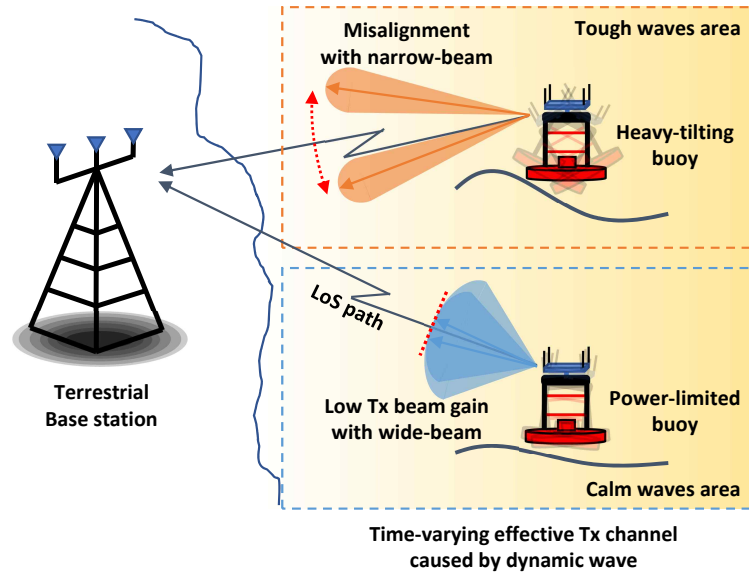


Figure 1. System model of the maritime buoy to base station uplink.

Statistical Modeling of Ocean Beam Channel

Incorporating the dynamic ocean environment into the channel model, the modified channel is expressed as

$$\mathbf{h}_{\text{dyn}} = \sqrt{\frac{K}{K+1}} \hat{\mathbf{h}}_{\text{LoS}} + \sqrt{\frac{1}{K+1}} \mathbf{h}_{\text{NLoS}}, \quad (4)$$

where

$$\hat{\mathbf{h}}_{\text{LoS}}[i] = e^{j\pi(i-1) \sin(\Theta_{\text{AoD}} + \Omega_0)} \quad (5)$$

where Θ_{AoD} is the angle of departure at the buoy and Ω_0 is derived from the buoy's angular displacement caused by ocean wave dynamics modeled via the ISSC spectrum [3]. To account for the impact of the beamwidth on the overall channel performance, the effective channel gain is defined as the projection of the channel vector onto the beamforming vector. Specifically, the effective channel gain H_G is expressed as follows:

$$H_G = \|\mathbf{h}_{\text{dyn}} \mathbf{f}\|^2, \quad (6)$$

where H_G denotes the effective channel gain after considering the effects of the dynamic ocean environment, and $\mathbf{f}$ represents the beamforming vector corresponding to a specific beamwidth [22]. In high sea states, large wave heights induce substantial angular variations in the buoy, thereby increasing the probability of steering misalignment, particularly when employing narrower transmit beams. While a narrower beam can provide larger array gain, its susceptibility to misalignment may severely degrade system performance under adverse conditions. Consequently, the buoy must adapt its beam vector to preserve robustness against wave-induced motion in the current sea state. The optimal beamwidths and the corresponding minimum transmit powers required to meet the target across different sea states are summarized in Table 1.

Table 1. Optimal beamwidth and minimum power required per sea state.

Sea State	Optimal Beamwidth (°)	Minimum Power (W)
1	2.81	0.0013
2	5.62	0.0031
3	11.25	0.0106
4	11.25	0.0304
5	22.50	0.0416
6	22.50	0.0563
7	22.50	0.0779
8	45.00	0.1373
9	45.00	0.1625

Note: Values are computed for a target rate of 6 bps/Hz.

3. Joint Beamwidth and Power Selection Using 2DTS

This section introduces a joint beamwidth and power selection strategy for maritime communication using 2DTS. In scenarios where the sea-state information is known at the buoy, effective adaptation is crucial. Unlike the Upper Confidence Bound (UCB) algorithm, which struggles under dynamic conditions due to its reliance on confidence intervals, TS achieves a better balance between exploration and exploitation through probabilistic sampling. The proposed approach extends standard TS by jointly selecting the best beamwidth and minimum transmit power, enabling network-assisted real-time adaptation to varying maritime environments. Simulation shows that 2DTS outperforms standard TS, leading to lower power consumption, higher reliability, and robust operation under dynamic and uncertain sea states.

3.1. Standard Thompson Sampling

Thompson Sampling (TS) is a popular Bayesian decision-making algorithm tailored for uncertain environments, particularly effective in MAB problems. The fundamental objective in a MAB setting is to maximize cumulative rewards by choosing among K possible actions (or arms), each with an unknown reward distribution. TS accomplishes this by sampling from posterior distributions of the reward means and updating these distributions as new data (rewards) are observed. Over time, it allocates resources to actions believed to yield higher returns while still periodically exploring other actions to gather new information. Below, we provide an expanded description of TS, emphasizing the Bayesian structure and adaptive update rules.

3.1.1. Bayesian Multi-Armed Bandit Formulation

Consider a MAB with K arms, where each arm a_k is associated with an unknown reward distribution. Let μ_k denote the (unknown) mean reward of arm k . The agent,

or decision-maker, seeks to maximize the cumulative reward obtained by sequentially selecting arms over T rounds. At each round t , the agent selects an arm a_{k^*} , observes a reward r_{k^*} drawn from the distribution with mean μ_{k^*} , and subsequently updates its posterior belief about μ_{k^*} based on the observed reward. The central challenge lies in balancing exploration of insufficiently understood arms with exploitation of arms already identified as yielding high rewards.

3.1.2. Bayesian Perspective and Thompson Sampling Logic

From a Bayesian standpoint, each mean reward μ_k is modeled with a prior distribution that reflects initial uncertainty. As rewards are observed, Bayes' rule is applied to update this belief, yielding a posterior distribution $P(\mu_k | \text{data})$. Thompson Sampling exploits this posterior by sampling a candidate value of μ_k for each arm and selecting the arm with the highest sampled value. This naturally balances exploration and exploitation, as arms with uncertain but potentially high means can still be chosen. The procedure can be described step by step as follows:

(1) Initialization.

Define a prior distribution $P(\mu_k)$ for each arm $k \in \{1, \dots, K\}$. The choice of prior depends on the reward type:

- For Bernoulli rewards (i.e., rewards are either 0 or 1), a beta distribution is typically selected:

$$\mu_k \sim \text{Beta}(\alpha_k, \beta_k), \quad (7)$$

where α_k and β_k are positive hyperparameters. This is because the beta distribution is the conjugate prior for the Bernoulli likelihood, ensuring a simple form of posterior updates.

- For Gaussian rewards, a normal prior is often chosen:

$$\mu_k \sim \mathcal{N}(\mu_{0,k}, \sigma_{0,k}^2), \quad (8)$$

where $\mu_{0,k}$ is the initial mean, and $\sigma_{0,k}^2$ is the initial variance. The normal distribution is a conjugate prior for the Gaussian likelihood, which keeps posterior updates tractable.

These priors capture the agent's initial beliefs about the reward means before any data is observed.

(2) Sampling.

In each round t , for each arm a_k , sample a reward estimate $\hat{\mu}_k$ from the posterior distribution $P(\mu_k | \text{data})$. This posterior reflects the agent's updated knowledge after previous selections. Concretely,

- Bernoulli rewards (Beta prior):

$$\hat{\mu}_k \sim \text{Beta}(\alpha_k + S_k, \beta_k + F_k), \quad (9)$$

where S_k is the total number of observed "successes" (reward = 1) and F_k the total number of observed "failures" (reward = 0) for arm k . Each time arm k is played, the parameters α_k, β_k are updated accordingly.

- Gaussian rewards (Normal prior):

$$\hat{\mu}_k \sim \mathcal{N}\left(\frac{\mu_{0,k}/\sigma_{0,k}^2 + \sum r_k/\sigma^2}{1/\sigma_{0,k}^2 + n_k/\sigma^2}, \frac{1}{1/\sigma_{0,k}^2 + n_k/\sigma^2}\right). \quad (10)$$

Here, $\sum r_k$ is the sum of all rewards observed from arm k , n_k is the number of times arm k has been selected, and σ^2 is the known variance of the underlying reward distribution.

By drawing a sample $\hat{\mu}_k$ from each posterior, TS naturally incorporates both the estimated mean and associated uncertainty.

(3) Action Selection.

After sampling one candidate $\hat{\mu}_k$ for each arm k , the algorithm chooses the arm

$$a_{k^*} = \arg \max_k \hat{\mu}_k. \quad (11)$$

Thus, the selected arm is the one that this particular sample indicates has the largest mean reward. Over many rounds, arms with higher posterior means or lower uncertainty will be chosen more often, but less-known arms still have a chance to be sampled if their posterior draws occasionally exceed those of well-explored arms.

(4) Update.

The chosen arm a_{k^*} is played, and the resulting reward r_{k^*} is observed. The posterior for μ_{k^*} is then updated according to Bayesian rules:

- Bernoulli rewards: If $r_{k^*} \in \{0, 1\}$, update

$$\alpha_{k^*} \leftarrow \alpha_{k^*} + r_{k^*}, \quad \beta_{k^*} \leftarrow \beta_{k^*} + (1 - r_{k^*}). \quad (12)$$

- Gaussian rewards: Update the mean and variance parameters of the normal posterior using standard Gaussian conjugate updating formulas, incorporating the new data point r_{k^*} .

This feedback loop is repeated in each subsequent round, enabling the algorithm to progressively refine its understanding of each arm's reward potential.

3.2. Problem Formulation

In this paper, the objective is to select the beamwidth and the transmit power level so that the achieved rate meets or exceeds a predefined threshold while minimizing the power used. To achieve this, we utilized 2DTS to select the best beamwidth and minimum power level.

Let θ_i denote the available beamwidth options, where $i \in 1, 2, \dots, N$, and P_j denote the available power levels, where $j \in 1, 2, \dots, M$. The objective is to maximize the cumulative number of successful transmissions over multiple trials while minimizing energy consumption.

3.3. Thompson Sampling with Joint Beamwidth and Power Selection

We apply 2DTS to jointly select the beamwidth and the power level in each trial to find the beamwidth and the minimum power that meet a given rate threshold r . The algorithm operates as follows:

- Initialize the parameters α and β for the beta distribution for each beamwidth i and each power level j , with $\alpha_i = 1$, $\beta_i = 1$, $\alpha_{i,j} = 1$, and $\beta_{i,j} = 1$ for all i and j .
- For each trial t
 - Sample θ_i from $\text{Beta}(\alpha_i, \beta_i)$ for each beamwidth i .
 - Select the beamwidth b_t with the highest sampled value:

$$b_t = \arg \max_i \theta_i \quad (13)$$

- For the selected beamwidth b_t , establish the power value corresponding to the maximum sampled probability from the beta distribution $P_{i,j}$ from $\text{Beta}(\alpha_{i,j}, \beta_{i,j})$ for each power level j , then select the power level p_t with the highest sampled value:

$$p_t = \arg \max_j P_{i,j} \quad (14)$$

The algorithm selects the power level p_t , aiming to minimize power usage by choosing the lowest possible power level that still satisfies the target threshold r . Here, R_t denotes the instantaneous channel capacity, while the threshold r is set strictly below this maximum capacity to represent the actual transmission rate. Thus, the algorithm checks whether the achievable rate R_t exceeds r , treating r as the success criterion for transmission.

- Calculate the SNR for the selected beamwidth and power level:

$$\text{SNR}_t = \frac{p_t H_G}{N}, \quad (15)$$

where H_G is the effective channel gain associated with beamwidth b_t and N is the noise power. Using this SNR, compute the rate R_t as follows:

$$R_t = \log_2(1 + \text{SNR}_t), \quad (16)$$

- Determine the reward r_t based on whether the achievable rate R_t meets or exceeds the target threshold r :

$$r_t = \begin{cases} 1, & \text{if } R_t \geq r \\ 0, & \text{otherwise} \end{cases} \quad (17)$$

- Update the beta distribution parameters for the selected beamwidth and power level based on the observed reward:

- * For the selected beamwidth b_t

$$\alpha_{b_t} = \alpha_{b_t} + r_t, \quad (18)$$

$$\beta_{b_t} = \beta_{b_t} + (1 - r_t), \quad (19)$$

- * For the selected power level p_t of the selected beamwidth b_t

- If $r_t = 1$, this means success (the rate threshold is met or exceeded); increase α for the selected power level and the level immediately below it (if it exists). This update encourages the selection of lower power levels that can meet the threshold, improving power efficiency:

$$\alpha_{b_t, p_t} = \alpha_{b_t, p_t} + r_t, \quad (20)$$

$$\alpha_{b_t, p_{t-1}} = \alpha_{b_t, p_{t-1}} + r_t \quad \text{if } p_{t-1} \text{ exists}, \quad (21)$$

- If $r_t = 0$, this means fail (the rate threshold is not met); increase β for the selected power level and all power levels below it. This penalizes lower power levels that fail to meet the threshold, encouraging the selection of higher power levels when necessary:

$$\beta_{b_t, 1:p_t} = \beta_{b_t, 1:p_t} + (1 - r_t), \quad (22)$$

In this approach, 2DTS iteratively updates its estimates of the success probabilities for each power level, allowing it to identify the minimum power level that achieves the required rate with high confidence. Adjusting α and β for each beamwidth and power level based on the observed rate, the algorithm efficiently balances exploration and exploitation, focusing on power levels that maximize energy efficiency while meeting the communication threshold Algorithm 1.

Algorithm 1: 2DTS for Beamwidth and Power selection

Require: Threshold r , Noise power N_0 , Beamwidth options θ , Power levels P , Rate R .

- 1: Initialize $\alpha_i = 1, \beta_i = 1$ for all $i \in b$, and $\alpha_{i,j} = 1, \beta_{i,j} = 1$ for all $i \in \theta, j \in P$.
 - 2: **for** $t = 1$ to T **do**
 - 3: Sample $\theta_i \sim \text{Beta}(\alpha_i, \beta_i)$ for each $i \in \theta$.
 - 4: Select beamwidth $b_t = \arg \max_i \theta_i$.
 - 5: For the selected b_t , sample $P_{i,j} \sim \text{Beta}(\alpha_{i,j}, \beta_{i,j})$ for each $j \in P$.
 - 6: Select power level $p_t = \arg \max_j P_{i,j}$.
 - 7: Compute SNR: $\frac{p_t H_g}{N}$.
 - 8: Compute rate: $R_t = \log_2(1 + \text{SNR}_t)$.
 - 9: Compute reward $r_t = \begin{cases} 1, & \text{if } R_t \geq r \\ 0, & \text{otherwise} \end{cases}$.
 - 10: **if** $r_t = 1$ **then**
 - 11: Update $\alpha_{b_t} \leftarrow \alpha_{b_t} + r_t$.
 - 12: Update $\alpha_{b_t, p_t} \leftarrow \alpha_{b_t, p_t} + r_t$.
 - 13: **if** p_{t-1} exists **then**
 - 14: Update $\alpha_{b_t, p_{t-1}} \leftarrow \alpha_{b_t, p_{t-1}} + r_t$.
 - 15: **end if**
 - 16: **else**
 - 17: Update $\beta_{b_t} \leftarrow \beta_{b_t} + r_t$.
 - 18: Update $\beta_{b_t, 1:p_t} \leftarrow \beta_{b_t, 1:p_t} + r_t$.
 - 19: **end if**
 - 20: **end for**
-

3.4. Energy Efficiency

We confirmed the proposed 2DTS joint beamwidth–power selection over T trials. At each trial t , 2DTS samples a beam and a power. The cumulative number of successful transmissions at trial t is

$$R_T = \sum_{t=1}^T r_t. \quad (23)$$

where $r_t \in \{0, 1\}$ denotes the binary reward at round t . The cumulative power used over T trials is

$$P_{\text{total}} = \sum_{t=1}^T p_t, \quad 0 < p_t \leq P_{\text{max}}, \quad b_t \in \theta, \quad p_t \in P. \quad (24)$$

Accordingly, the overall energy efficiency after T trials is

$$\text{EE}_T = \frac{R_T}{P_{\text{total}}} = \frac{\sum_{t=1}^T r_t}{\sum_{t=1}^T p_t}. \quad (25)$$

Equivalently, 2DTS seeks to maximize EE_T over the following sequence:

$$\begin{aligned} & \underset{\{(b_t, p_t)\}_{t=1}^T}{\text{maximize}} \quad EE_T = \frac{R_t}{P_{\text{total}}} \\ & \text{subject to} \quad r_t = \begin{cases} 1, & \log_2(1 + \text{SNR}_t) \geq r, \\ 0, & \text{otherwise,} \end{cases} \\ & \quad 0 < p_t \leq p_{\max}, \quad b_t \in \theta, \quad p_t \in P, \quad t = 1, \dots, T. \end{aligned}$$

The results indicate that the algorithm successfully converges to the optimal beamwidth and power level, achieving reliable communication with minimum energy consumption.

4. Simulation Results

In this section, we present simulation results that quantitatively demonstrate the effectiveness of the proposed 2DTS algorithm in selecting the optimal beamwidth and transmit power under varying sea states. Performance is measured in terms of cumulative energy efficiency (r/p) in iterative rounds. Here, r represents the achieved data rate (in bps/Hz) obtained from the simulation, rather than the binary success indicator used within the Thompson Sampling process. Accordingly, the energy efficiency (EE) metric has the physical unit of bits/Joule/Hz, consistent with the standard definition of spectral energy efficiency. For clarity, all relevant figures are labeled as “Energy Efficiency (bps/Hz/J)”. In buoy-based maritime uplinks, energy efficiency (EE) is a primary design objective due to strict power budgets. Throughout this section, we use $EE = r/P_t$, where r denotes the achieved rate in bps/Hz (i.e., spectral efficiency, SE). Hence, EE implicitly captures SE in its numerator while simultaneously reflecting the rate–power trade-off that is central to battery-constrained buoy operation. In our target-rate setting, methods that satisfy the reliability constraint yield comparable SE at the threshold, so the observed EE gains of 2DTS primarily arise from using lower transmit power to meet the same target. Qualitatively, therefore, SE trends are consistent with the reported EE improvements. Figures 2 and 3 show the energy efficiency achieved by four schemes: (i) an ideal case that assumes perfect prior knowledge of the optimal beamwidth and power at each round and (ii) the proposed 2DTS algorithm. (iii) The “modified TS with a dedicated exploration phase” (Exploration TS) follows an explore–then–commit protocol over the joint action space of beamwidth and power. During the first T_{exp} rounds, actions are sampled uniformly (no exploitation) and beta posteriors are updated from binary success/failure. At $t = T_{\text{exp}} + 1$, the algorithm commits to a single pair: among arms whose posterior-mean success probability meets the target reliability, it picks the minimum-power configuration; if none is feasible, it selects the arm with the largest posterior mean and uses it for the remaining rounds and (iv) a conventional TS that independently selects beamwidth and power using standard TS. As shown in Figure 2, under a moderate sea state (sea state 4), the proposed 2DTS achieves an energy efficiency of approximately 1.26×10^4 (bps/Hz/J) at round 600, which is 73.5% of the ideal case performance (1.71×10^4 (bps/Hz/J)). This result clearly outperforms both the exploration-based variant (0.64×10^4) and the conventional TS (0.23×10^4), indicating relative gains of 96.9% and 447.8% over the exploration-based and conventional TS base-lines, respectively. As shown in Figure 3, for a harsh sea state (sea state 9), the performance gap narrows further. In round 600, 2DTS achieves 3.09×10^4 (bps/Hz/J), corresponding to 85.6% of the ideal case (3.61×10^4), while exploration TS and conventional TS reach 1.68×10^4 and 1.45×10^4 , respectively. This translates to a performance gain of 84.0% over conventional TS, again highlighting the robustness of the 2DTS framework in dynamically

fluctuating maritime environments. The superior performance of 2DTS arises from its joint optimization structure.

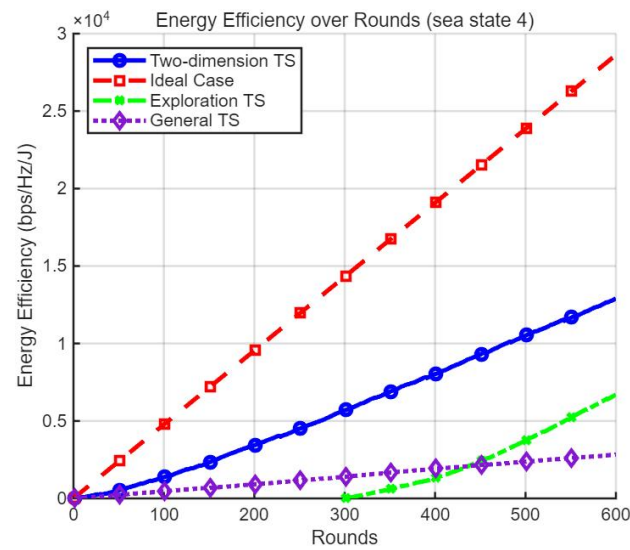


Figure 2. Energy efficiency over rounds (sea state 4).

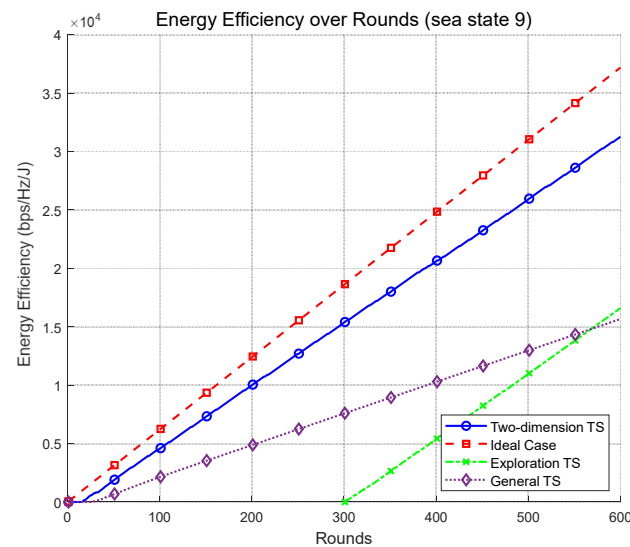


Figure 3. Energy efficiency over rounds (sea state 9).

5. Discussion

5.1. Key Findings

Across moderate (Sea State 4) and harsh (Sea State 9) sea states, the proposed 2DTS consistently achieves high spectral energy efficiency (EE) while meeting the target-rate constraint. At round 600, 2DTS attains 1.26×10^4 bps/Hz/J in sea state 4 (73.5% of the ideal) and 3.09×10^4 bps/Hz/J in sea state 9 (85.6% of the ideal), outperforming exploration-then-commit TS and conventional TS by large margins. These results indicate that feasibility-first selection combined with minimum-power control is effective under sea-state uncertainty.

5.2. Joint vs. Independent Selection

Unlike conventional TS that treats beamwidth and power independently, 2DTS jointly updates and selects (beamwidth, power) pairs. This joint treatment captures the coupling between directional gain (beamwidth) and energy usage (power), reducing both

misalignment-induced losses and unnecessary transmit power. The observed gains over conventional TS in Figures 2 and 3 are consistent with this mechanism.

5.3. Effect of Explore–Then–Commit

The explore–then–commit baseline improves naive independent selection by dedicating an initial exploration budget; however, its one-shot commitment is vulnerable to subsequent sea-state shifts. In contrast, 2DTS continuously re-balances exploration and exploitation, which leads to higher cumulative EE across rounds.

5.4. Sensitivity and Limitations

Performance depends on the reliability threshold, the discretization of the (beamwidth, power) action grid, and the duration of the exploratory budget in the explore–then–commit baseline (T_{exp}). While our setup requires only binary success/failure feedback (compatible with LTE-M systems lacking full MIMO/Channel State Information), extremely severe sea states or overly coarse action grids may slow convergence. Incorporating coarse contextual signals (e.g., forecasted sea state) or adaptive action refinement could further improve sample efficiency.

5.5. Practical Implications and Future Work

The results suggest that lightweight, feedback-driven learning can provide reliable and energy-efficient operation for buoy-based uplinks without explicit sea-state information. Future work includes context-aware 2DTS (e.g., integrating satellite-based forecasts as side information), multi-buoy cooperation, and relay-assisted links.

Summary and Novelty

Unlike standard TS, which treats each parameter selection problem independently, our 2DTS algorithm simultaneously selects the optimal beamwidth–power pair through a joint posterior update that captures both directional-gain dynamics (via beamwidth) and power–efficiency trade-offs. This design is tailored to maritime systems where beam misalignment and power constraints are tightly coupled due to wave-induced platform motion. To the best of our knowledge, no existing TS-based algorithm directly addresses the joint selection of beamwidth and power levels under unknown and dynamic sea-state conditions. Although advanced variants of TS (e.g., contextual or neural TS) have been proposed in other domains, they are generally ill-suited to the maritime scenario considered here. In our setting, the decision space is discrete and multidimensional and is tightly coupled with physical constraints such as misalignment and energy efficiency. Therefore, our approach constitutes a novel TS formulation tailored to this setting, offering a practical solution when sea-state information is unavailable. In summary, 2DTS achieves up to 85.6% of the ideal energy efficiency across varying sea states while significantly outperforming baseline methods, demonstrating its suitability for energy-constrained maritime communication environments.

6. Conclusions

In this work, we proposed an adaptive approach for buoy-embedded maritime communication systems, focusing on the selection of beamforming and power control under dynamic sea conditions. For unknown sea states, we developed a novel 2DTS algorithm that simultaneously adjusts beamwidth and transmission power, achieving robust communication even without direct sea-state information. This adaptive framework balances exploration and exploitation to optimize communication parameters over time, closely approximating ideal performance and ensuring energy-efficient operation. Our simulation

results validate the effectiveness of this method, demonstrating significant improvements in energy efficiency and reliability in dynamic maritime environments. Our approach advances the field of maritime communications by providing a robust, energy-efficient strategy for dynamic and unpredictable sea conditions.

Author Contributions: Conceptualization, K.J.L. and D.K.; methodology, K.J.L.; software, K.J.L.; validation, J.-H.J. and D.K.; writing—original draft, K.J.L.; writing—review and editing, J.-H.J., S.C., K.-W.K., and D.K.; supervision, D.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Korea Institute of Marine Science and Technology Promotion (KIMST), funded by the Ministry of Oceans and Fisheries, under grant number RS-2021-KS211516.

Data Availability Statement: Data supporting the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments: The authors would like to thank the Korea Institute of Marine Science and Technology Promotion (KIMST) and the Ministry of Oceans and Fisheries for their continued support under the project RS-2023-00238653. The authors also appreciate all collaborators for valuable discussions and data sharing.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Palma, D. Enabling the Maritime Internet of Things: CoAP and 6LoWPAN Performance Over VHF Links. *IEEE Internet Things J.* **2018**, *5*, 5205–5212. [CrossRef]
- Yang, T.; Zheng, Z.; Liang, H.; Deng, R.; Cheng, N.; Shen, X. Green Energy and Content-Aware Data Transmissions in Maritime Wireless Communication Networks. *IEEE Trans. Intell. Transp. Syst.* **2015**, *16*, 751–762. [CrossRef]
- Huo, Y.; Dong, X.; Beatty, S. Cellular Communications in Ocean Waves for Maritime Internet of Things. *IEEE Internet Things J.* **2020**, *7*, 9965–9979. [CrossRef]
- Zhou, Z.; Ge, N.; Wang, Z. Two-Timescale Beam Selection and Power Allocation for Maritime Offshore Communications. *IEEE Commun. Lett.* **2021**, *25*, 3060–3064. [CrossRef]
- Guan, S.; Wang, J.; Jiang, C.; Duan, R.; Ren, Y.; Quek, T.Q.S. MagicNet: The Maritime Giant Cellular Network. *IEEE Commun. Mag.* **2021**, *59*, 117–123. [CrossRef]
- Kim, H.J.; Tiwari, S.V.; Chung, Y.H. Multi-hop relay-based maritime visible light communication. *Chin. Opt. Lett.* **2016**, *14*, 050607. [CrossRef]
- Wang, W.; Gill, E.W. Evaluation of Beamforming and Direction Finding for a Phased Array HF Ocean Current Radar. *J. Atmos. Ocean. Technol.* **2016**, *33*, 2599–2613. [CrossRef]
- Romdhane, I.; Kaddoum, G. A Reinforcement-Learning-Based Beam Adaptation for Underwater Optical Wireless Communications. *IEEE Internet Things J.* **2022**, *9*, 20270–20281. [CrossRef]
- Jo, S.W.; Shim, W.S. LTE-Maritime: High-Speed Maritime Wireless Communication Based on LTE Technology. *IEEE Access* **2019**, *7*, 53172–53181. [CrossRef]
- Zhang, L.; Wang, Z.; Zhang, H.; Min, M.; Wang, C.; Niyato, D.; Han, Z. Anti-Jamming Colonel Blotto Game for Underwater Acoustic Backscatter Communication. *IEEE Trans. Veh. Technol.* **2024**, *73*, 10181–10195. [CrossRef]
- Ibrahim, S.; Mostafa, M.; Jnadi, A.; Salloum, H.; Osinenko, P. Comprehensive Overview of Reward Engineering and Shaping in Advancing Reinforcement Learning Applications. *IEEE Access* **2024**, *12*, 175473–175500. [CrossRef]
- Russo, D.J.; Roy, B.V.; Kazerouni, A.; Osband, I.; Wen, Z. A Tutorial on Thompson Sampling. *Found. Trends Mach. Learn.* **2018**, *11*, 1–96. [CrossRef]
- Chapelle, O.; Li, L. An Empirical Evaluation of Thompson Sampling. In *Proceedings of the Advances in Neural Information Processing Systems*; Shawe-Taylor, J., Zemel, R., Bartlett, P., Pereira, F., Weinberger, K., Eds.; Curran Associates, Inc.: San Francisco, CA, USA, 2011; Volume 24. Available online: https://proceedings.neurips.cc/paper_files/paper/2011/file/e53a0a2978c28872a4505bdb51db06dc-Paper.pdf (accessed on 12 December 2011).
- Deng, W.; Kamiya, S.; Yamamoto, K.; Nishio, T.; Morikura, M. Thompson Sampling-Based Channel Selection Through Density Estimation Aided by Stochastic Geometry. *IEEE Access* **2020**, *8*, 14841–14850. [CrossRef]

15. Tong, J.; Fu, L.; Wang, Y.; Han, Z. Model-Based Thompson Sampling for Frequency and Rate Selection in Underwater Acoustic Communications. *IEEE Trans. Wireless Commun.* **2023**, *22*, 6946–6961. [[CrossRef](#)]
16. Komiyama, J.; Honda, J.; Nakagawa, H. Optimal Regret Analysis of Thompson Sampling in Stochastic Multi-armed Bandit Problem with Multiple Plays. In Proceedings of the 32nd International Conference on Machine Learning, Lille, France, 7–9 July 2015; Volume 37, pp. 1152–1161. [[CrossRef](#)]
17. Bai, L.; Han, R.; Liu, J.; Choi, J.; Zhang, W. Random Access and Detection Performance of Internet of Things for Smart Ocean. *IEEE Internet Things J.* **2020**, *7*, 9858–9869. [[CrossRef](#)]
18. Goodwin, M.; Elko, G. Constant beamwidth beamforming. In Proceedings of the 1993 IEEE International Conference on Acoustics, Speech, and Signal Processing, Minneapolis, MN, USA, 27–30 April 1993; Volume 1, pp. 169–172. [[CrossRef](#)]
19. Duan, R.; Wang, J.; Zhang, H.; Ren, Y.; Hanzo, L. Joint Multicast Beamforming and Relay Design for Maritime Communication Systems. *IEEE Trans. Green Commun. Netw.* **2020**, *4*, 139–151. [[CrossRef](#)]
20. Wang, J.; Zhou, H.; Li, Y.; Sun, Q.; Wu, Y.; Jin, S.; Quek, T.Q.S.; Xu, C. Wireless Channel Models for Maritime Communications. *IEEE Access* **2018**, *6*, 68070–68088. [[CrossRef](#)]
21. Yau, K.L.A.; Syed, A.R.; Hashim, W.; Qadir, J.; Wu, C.; Hassan, N. Maritime Networking: Bringing Internet to the Sea. *IEEE Access* **2019**, *7*, 48236–48255. [[CrossRef](#)]
22. Love, D.; Heath, R. Grassmannian beamforming on correlated MIMO channels. In Proceedings of the IEEE Global Telecommunications Conference, GLOBECOM '04, Dallas, TX, USA, 29 November–3 December 2004; Volume 1, pp. 106–110. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.