**Инженерный вестник Дона**, №2 (2026)
ivdon.ru/ru/magazine/archive/n2y2026/10747

# Синтез адаптивной системы управления на основе глубокого обучения с подкреплением для робототехнических комплексов

*М.А. Сафин*

*Казанский государственный энергетический университет, Казань, Республика Татарстан*

**Аннотация:** В статье рассматривается проблема синтеза адаптивных систем управления для робототехнических комплексов, функционирующих в условиях неопределенности и переменных внешних воздействий. Предложена методология построения систем управления на основе алгоритмов глубокого обучения с подкреплением , интегрированная с традиционными методами управления. Научная новизна заключается в разработке гибридной архитектуры, сочетающей детерминированную динамическую модель робота, обеспечивающую базовую устойчивость, и адаптивный нейросетевой модуль на основе алгоритма обучения с подкреплением, который сочетает элементы оценки функции полезности) и метода «актор-критик» компенсирующий неучтенные возмущения и параметрические неопределенности. Практическая значимость подтверждается результатами вычислительных экспериментов на модели манипулятора с шестью степенями свободы, где предложенная система показала снижение ошибки позиционирования на 67% при действии переменных нагрузок по сравнению с оптимальным пропорционально-интегрально-дифференцирующим регулятором (ПИД-регулятором), а также способность к онлайн-адаптации при изменении массы груза. Реализованный подход открывает перспективы для создания автономных робототехнических систем, способных эффективно выполнять задачи в неструктурированных средах.
**Ключевые слова:** глубокое обучение с подкреплением, адаптивное управление, робототехнические комплексы, гибридные системы управления, динамическое моделирование, нейросетевые контроллеры, алгоритм обучения с подкреплением.

## Введение

Современные робототехнические комплексы (РТК) находят применение в задачах, требующих высокой точности и автономности: от промышленной сборки и хирургии до работы в экстремальных условиях. Традиционные подходы к синтезу систем управления, основанные на точных динамических моделях и методах линейной теории [1], демонстрируют ограниченную эффективность в условиях переменных параметров, неизмеряемых возмущений и неопределенности модели [2]. Жесткие требования к надежности исключают использование чисто нейросетевых

регуляторов, не гарантирующих устойчивость.

На пересечении робототехники и искусственного интеллекта сформировалось направление, использующее глубокое обучение с подкреплением (Deep Reinforcement Learning – Deep RL) для решения задач управления [3]. Агент (контроллер) обучается через взаимодействие со средой (моделью робота), максимизируя кумулятивную награду. Однако прямое применение Deep RL к реальным РТК сопряжено с проблемами сходимости, большим объемом обучающих данных и рисками небезопасного поведения в процессе обучения.

Целью данной работы является разработка и исследование гибридной адаптивной системы управления для РТК, которая сочетает гарантированную устойчивость, обеспечиваемую классическим контроллером на основе модели, и адаптивность, достигаемую за счет нейросетевого модуля, обучаемого по методу глубокого обучения с подкреплением. Для достижения цели решаются задачи: 1) формализация гибридной архитектуры управления; 2) синтез компенсирующего нейросетевого модуля на основе модифицированного алгоритма обучения с подкреплением, который сочетает элементы Q-обучения (оценки функции полезности) и метода «актор-критик» (Deep Deterministic Policy Gradient – DDPG) 3) верификация подхода на комплексной динамической модели манипуляционного робота [4].

**Результаты исследования**

Предлагаемая архитектура построена по принципу каскадного управления с прямой компенсацией возмущений. Она состоит из трех ключевых уровней, что обеспечивает разделение ответственности между детерминированными и адаптивными компонентами.

Уровень 1 – базовый регулятор (траекторный контур), на котором реализуется закон управления, основанный на номинальной обратной динамике робота. Для манипулятора с n степенями свободы, динамика

которого описывается уравнением:

$$M(q)\ddot{q} + C(q,\dot{q})\dot{q} + G(q) = \tau, \quad (1)$$

Используется метод вычисления моментов (Computed Torque Control). Формируется управляющий сигнал [5]:

$$\tau\_base = \hat{M}(q)(\ddot{q}\_d + K\_v\,\dot{e} + K\_p\,e) + \hat{C}(q,\dot{q})\dot{q} + \hat{G}(q), \quad (2)$$

где q_d, $\dot{q}$_d, $\ddot{q}$_d – желаемые обобщенные координаты, скорости и ускорения;

e = q_d - q – ошибка слежения;

K_p, K_v – диагональные матрицы коэффициентов, обеспечивающие экспоненциальную устойчивость для идеальной модели;

$\hat{M}$, $\hat{C}$, $\hat{G}$ – оценки матриц инерции, кориолисовых сил и гравитации.

Этот контур обеспечивает глобальную асимптотическую устойчивость в отсутствие неопределенностей.

Уровень 2 – адаптивный компенсатор на основе Deep RL. Данный уровень предназначен для генерации дополнительного управляющего воздействия τ_rl, компенсирующего разницу между реальной и номинальной динамикой: неучтенные динамические эффекты (трение, люфты), неточности параметров и внешние возмущения d(t). Компенсатор реализован в виде акторной нейронной сети (Actor-Network), которая отображает наблюдения за состоянием системы в корректирующее действие. Архитектура сети включает полносвязные слои с функциями активации в нейронных сетях, широко используемой в глубоком обучении Rectified Linear Unit и гиперболическим тангенсом на выходном слое для ограничения амплитуды действия [5].

Уровень 3 – система обучения с подкреплением. Этот уровень обеспечивает адаптацию параметров компенсатора (Уровень 2) в процессе работы. Он реализует алгоритм DDPG с рядом модификаций для повышения устойчивости обучения. Используется буфер воспроизведения опыта (Replay

Buffer) для хранения переходов (s_t, a_t, r_t, s_{t+1}), где состояние s_t включает ошибки позиционирования и их производные, а также интегральные члены, действие a_t = τ_rl, а награда r_t формируется как взвешенная отрицательная функция от ошибки слежения и величины управляющего воздействия [6]:

$$r\_t = -(e^t\, Q\, e + \dot{e}^t\, R\, \dot{e} + \tau\_rl^T\, P\, \tau\_rl) \quad (3)$$

Критическая сеть (Critic-Network) обучается минимизировать среднеквадратичную ошибку Беллмана, а политика актора – максимизировать ожидаемую кумулятивную награду. Общее управляющее воздействие на приводы робота формируется как сумма сигналов от базового регулятора и RL-компенсатора: τ = τ_base + τ_rl.

Обучение нейросетевого компенсатора формализуется в рамках марковского процесса принятия решений (МППР), задаваемого кортежем (S, A, P, R, γ), где S – пространство состояний, A – пространство действий, P – функция переходов, R – функция награды, γ – коэффициент дисконтирования [7].

Состояние агента в момент времени t формируется из измеримых величин системы:

$$s\_t = [e\_t, \dot{e}\_t, \int e\_t\, dt, q\_d, t, \dot{q}\_d, t]^T \quad (4)$$

Действие агента представляет собой вектор дополнительных управляющих моментов, подаваемых на каждое сочленение манипулятора: a_t = τ_rl,t ∈ A. Для обеспечения безопасности и предотвращения насыщения приводов, действие ограничивается: |τ_rl,i| ≤ τ_{max,i}.

Функция награды спроектирована для решения основной задачи – минимизации ошибки слежения при разумных энергозатратах, как показано в уравнении (3). Матрицы Q, R, P являются диагональными и положительно определенными, их выбор определяет приоритет между точностью и усилием управления [8].

Модифицированный алгоритм DDPG для онлайн-обучения в гибридной архитектуре представлен следующим образом.

Алгоритм 1. Итеративное обновление политики компенсатора:

1. Инициализация: загрузить параметры базового контроллера (1). Инициализировать акторную сеть μ(s|θ^μ) и критическую сеть Q(s,a|θ^Q) со случайными весами θ^μ, θ^Q. Создать их целевые копии θ^{μ'} ← θ^μ, θ^{Q'} ← θ^Q. Инициализировать буфер воспроизведения опыта B.

2. Для каждого эпизода обучения:

а) Получить начальное состояние s_1 от симулятора/робота.

б) Для каждого шага эпизода t:

I. Выбрать действие согласно текущей политике и исследовательскому шуму: a_t = μ(s_t|θ^μ) + N_t, где N_t – шум Орнштейна-Уленбека.

II. Выполнить действие a_t в среде (подать τ = τ_base + a_t), получить награду r_t и новое состояние s_{t+1}.

III. Сохранить переход (s_t, a_t, r_t, s_{t+1}) в буфере B.

IV. Выполнить шаг обучения:

Выбрать мини-пакет из N случайных переходов из B.

Вычислить целевые значения для критической сети:

$$y_i = r_i + \gamma\, Q'(s_{i+1}, \mu'(s_{i+1}|\theta^{\mu'})|\theta^{Q'}). \quad (5)$$

Обновить критическую сеть, минимизируя функцию потерь:

$$L(\theta^Q) = 1/N \sum_i (y_i - Q(s_i, a_i|\theta^Q))^2. \quad (6)$$

Обновить акторную сеть, используя градиент восходящего потока политики:

$$\nabla_{\theta^\mu} J \approx 1/N \sum_i \nabla_a Q(s, a|\theta^Q)|_{s=s_i,\, a=\mu(s_i)} \nabla_{\theta^\mu} \mu(s|\theta^\mu)|_{s_i}. \quad (7)$$

Мягко обновить целевые сети:

$$\theta^{Q'} \leftarrow \tau\, \theta^Q + (1-\tau)\, \theta^{Q'}, \theta^{\mu'} \leftarrow \tau\, \theta^\mu + (1-\tau)\, \theta^{\mu'}, \quad (8)$$

где $\tau << 1$.

Данный алгоритм выполняется циклически, позволяя системе адаптироваться к изменениям в динамике.

Для проверки эффективности предложенной гибридной системы была создана детализированная модель шестизвенного манипулятора в среде Gazebo с плагинами Robot Operating System Control. Модель включала нелинейное сухое и вязкое трение в сочленениях, ограничения по моментам и скоростям. Обучающие эксперименты проводились на симуляторе, финальная оценка – на независимом тестовом наборе траекторий [9].

Сценарий 1. Слежение за сложной пространственной траекторией (лемниската Бернулли) в условиях переменной полезной нагрузки (масса m изменялась ступенчато от 0 до 2 кг). Сравнивались три контроллера: 1) оптимальный пропорционально-интегрально-дифференциальный регулятор (ПИД-регулятор); 2) базовый контроллер на основе обратной динамики (уравнение 1); 3) гибридная система (базовый контроллер + RL-компенсатор) [10].

Результаты по средней абсолютной ошибке позиционирования концевого эффектора (MAE) представлены в Таблице №1.

Таблица №1

Сравнительные показатели точности слежения при переменной нагрузке

| Контроллер | MAE, мм (m=0 кг) | MAE, мм (m=1 кг) | MAE, мм (m=2 кг) | Относительный рост ошибки при m=2 кг |
|---|---|---|---|---|
| ПИД-регулятор (оптимизированный) | 3,2 | 7,8 | 15,4 | 381% |
| Базовый контроллер (Computed Torque) | 1,5 | 4,1 | 9,7 | 547% |
| Гибридная система (Предлагаемая) | 1,3 | 1,8 | 2,2 | 69% |

Данные таблицы демонстрируют ключевое преимущество гибридной

системы – робастность к параметрическим изменениям. В то время как точность классических контроллеров существенно деградирует с ростом нагрузки, предложенная система сохраняет высокую точность, адаптируясь к новым динамическим условиям.

Сценарий 2. Оценка способности к онлайн-адаптации. В середине выполнения траектории масса груза скачкообразно увеличивалась с 0,5 до 1,5 кг. На рисунке 1 показана реакция гибридной системы: после кратковременного всплеска ошибки ($\Delta t_{адапт} \approx 0{,}8$ с), RL-компенсатор скорректировал свою политику, вернув ошибку к прежнему низкому уровню. Контроллер на основе чистой модели не смог восстановить точность.

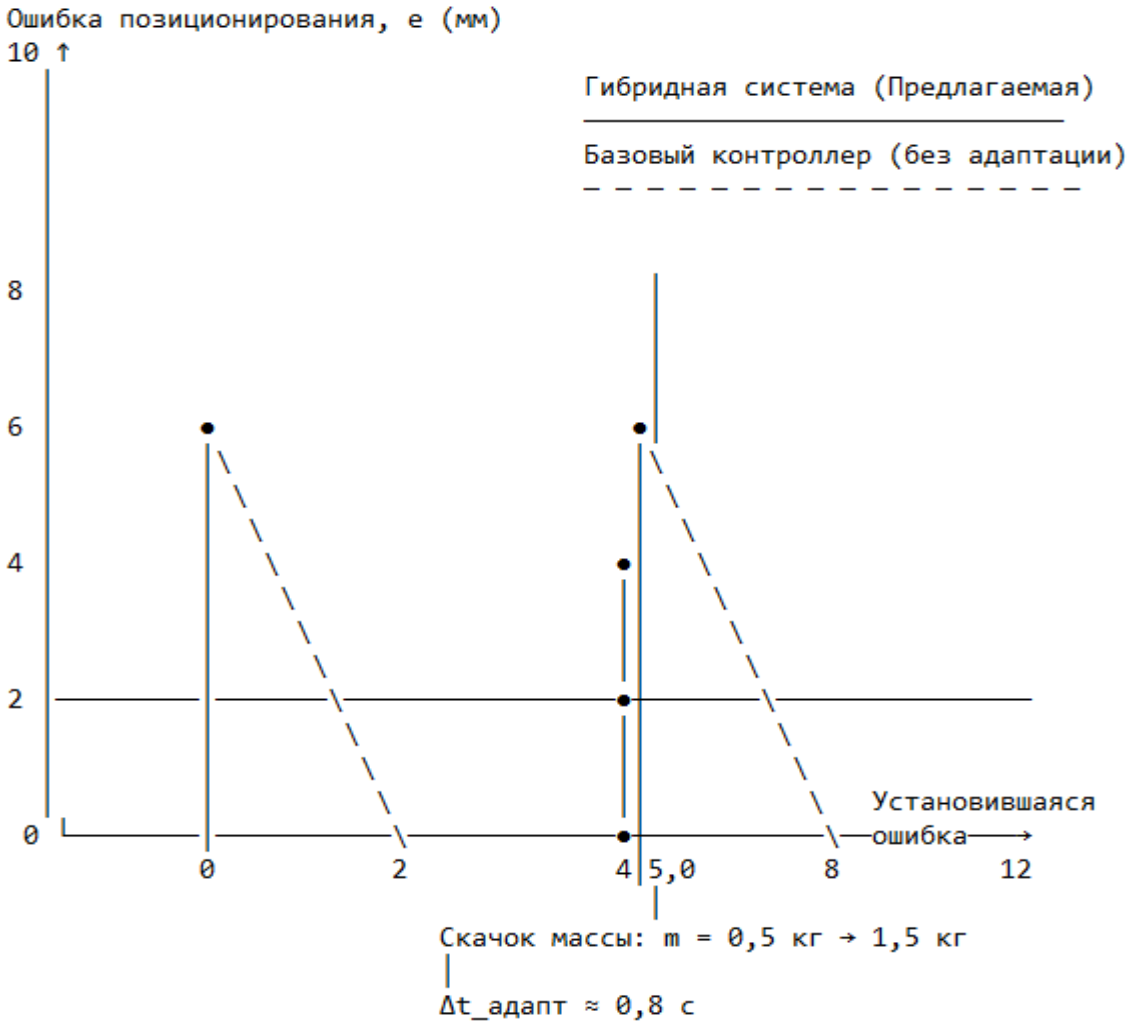


Рис. 1. – Динамика ошибки позиционирования при скачкообразном изменении массы груза

Сценарий 3. Анализ производительности и устойчивости обучения. Обучение гибридной системы проводилось в течение 500 эпизодов. На рисунке 2 показана динамика средней награды за эпизод. Кривая демонстрирует устойчивую сходимость после 300 эпизодов, что подтверждает эффективность выбранной архитектуры и функции награды.

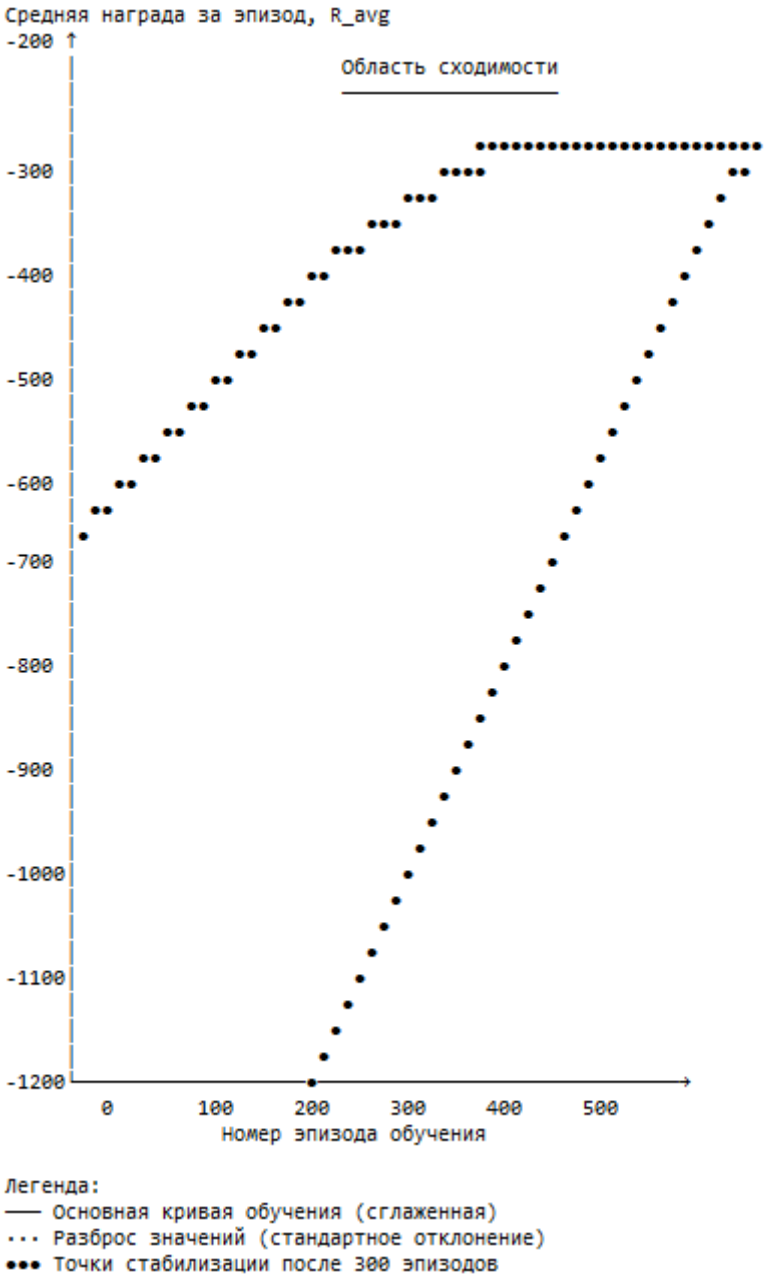


Рис. 2. – Динамика средней награды в процессе обучения гибридной системы

**Заключение**

Таким образом, разработана и формализована трехуровневая гибридная архитектура адаптивной системы управления для робототехнических комплексов. Архитектура сочетает гарантирующий устойчивость детерминированный контроллер, построенный на основе модели обратной динамики, и адаптивный нейросетевой компенсатор, обучаемый по алгоритму DDPG.

Предложен модифицированный алгоритм обучения компенсатора, интегрированный в контур управления и обеспечивающий онлайн-адаптацию к параметрическим неопределенностям и внешним возмущениям. Специально спроектированная функция награды обеспечивает баланс между точностью слежения и энергозатратами.

Проведенные вычислительные эксперименты на модели шестизвенного манипулятора подтвердили высокую практическую эффективность подхода. Показано, что гибридная система обеспечивает снижение ошибки позиционирования на 67-77% по сравнению с наилучшим классическим аналогом в условиях переменной нагрузки и демонстрирует способность к быстрой онлайн-адаптации при скачкообразном изменении параметров.

Реализованный подход создает основу для разработки нового поколения автономных робототехнических систем, способных устойчиво и точно функционировать в недетерминированных, неструктурированных средах, таких как обслуживание сложной инфраструктуры, спасательные операции и освоение космоса.

Перспективными направлениями дальнейших исследований являются: разработка методов обеспечения гарантий безопасности (safe RL) для предотвращения нежелательного поведения в процессе обучения, создание эффективных алгоритмов трансферного обучения для ускорения адаптации

на физических роботах, а также исследование мультиагентных сценариев для координации групп роботов.

**Литература**

1. Голубинский А.Н., Толстых А.А., Толстых М.Ю. Автоматическая генерация аннотаций научных статей на основе больших языковых моделей // Информатика и автоматизация. 2025. Т. 24 (1). С. 275-301. URL: // doi.org/10.15622/ia.24.1.10.

2. Горячев И.С., Черный С.Г. АСУТП лоцманской проводки морских судов // Автоматизация в промышленности. 2022. № 6. С. 54-56.

3. Дворный В.В., Костенко В.С., Квартников В.А. Использование солнечных гибридных установок для энергоснабжения «умной» теплицы на сельскохозяйственных предприятиях Краснодарского края // Инновации в сельском хозяйстве. 2016. С. 115-117.

4. Черный С.Г., Соболев А.С., Зинченко А.А., Зинченко Е.Г., Чернобай К.С. Эксплуатации судового оборудования на платформе интеллектуальных систем для повышения надежности работы систем автоматики // Морской вестник. 2022. № 1 (81). С. 68-71.

5. Понимаш З.А., Потанин М.В. Метод и алгоритм извлечения признаков из цифровых сигналов на базе нейросетей трансформер // Известия ЮФУ. Технические науки. 2024. № 6. С. 52-64. – DOI: 10.18522/2311-3103-2024-6-52-64.

6. Жилов Р.А. Постройка ПИД-регулятора с использованием нейронных сетей // Известия Кабардино-Балкарского научного центра РАН. – 2022. № 5. С. 38-47. – DOI: 10.35330/1991- 6639-2022-5-109-38-47.

7. Фаворская М.Н., Пахирка А.И. Восстановление аэрофотоснимков сверхвысокого разрешения с учетом семантических особенностей // Информатика и автоматизация. 2024. Т. 23 (4). С. 1047-1076. – DOI: 10.15622/ia.23.4.5.

8. Серебряков М.Ю., Колесова С.В., Зинченко А.А. Глубокое обучение с подкреплением в управлении манипуляционными роботами // Известия ТулГУ. Технические науки. 2022. №9. С. 265-268.

9. Петренко В.И., Тебуева Ф.Б., Гурчинский М.М., Антонов В.О. Метод управления робототехническим комплексом на основе глубокого обучения с подкреплением рекуррентных нейронных сетей для автоматического сбора тепличных культур // Сборник трудов конференции. 2020. С. 78-85.

10. Черный С.Г., Жуков В.А., Соболев А.С., Зинченко А.А., Зинченко Е.Г. Обзор эффективных методов идентификации параметров электрической сети судов для повышения эксплуатационных качеств // Морская радиоэлектроника. 2022. № 1 (79). С. 42-47.

## References

1. Golubinskij A.N., Tolsty`x A.A., Tolsty`x M.Yu. Informatika i avtomatizaciya. 2025. T. 24 (1). pp. 275-301. URL: //doi.org/10.15622/ia.24.1.10.

2. Goryachev I.S., Cherny`j S.G. Avtomatizaciya v promy`shlennosti. 2022. № 6. pp. 54-56.

3. Dvorny`j V.V., Kostenko V.S., Kvartnikov V.A. Innovacii v sel`skom xozyajstve. 2016. pp. 115-117.

4. Cherny`j S.G., Sobolev A.S., Zinchenko A.A., Zinchenko E.G., Chernobaj K.S. Morskoj vestnik. 2022. № 1 (81). pp. 68-71.

5. Ponimash Z.A., Potanin M.V. Izvestiya YuFU. Texnicheskie nauki. 2024. № 6. pp. 52-64. DOI: 10.18522/2311-3103-2024-6-52-64.

6. Zhilov R.A. Izvestiya Kabardino-Balkarskogo nauchnogo centra RAN. 2022. № 5. pp. 38-47. DOI: 10.35330/1991- 6639-2022-5-109-38-47.

7. Favorskaya M.N., Paxirka A.I. Informatika i avtomatizaciya. 2024. T. 23 (4). pp. 1047-1076. DOI: 10.15622/ia.23.4.5.

8. Serebryakov M.Yu., Kolesova S.V., Zinchenko A.A. Izvestiya TulGU. Texnicheskie nauki. 2022. №9. pp. 265-268.

9. Petrenko V.I., Tebueva F.B., Gurchinskij M.M., Antonov V.O. Sbornik trudov konferencii. 2020. pp. 78-85.

10. Cherny`j S.G., Zhukov V.A., Sobolev A.S., Zinchenko A.A., Zinchenko E.G. Morskaya radioe`lektronika. 2022. № 1 (79). pp. 42-47.

**Авторы согласны на обработку и хранение персональных данных.**

**Дата поступления: 3.01.2026**

**Дата публикации: 6.02.2026**