MAQR：面向多跳生物纳米物联网新鲜状态传递的轻量级记忆感知Q路由

Ruifeng Zheng*、Veronika Volkova*、Zhihan Xu*、Pengjie Zhou*‡、Mart´ın Schottlender*、Rafael F. Schaefer†‡、Frank H.P. Fitzek*‡和Pit Hofmann*‡

*德国电信通信网络教席，德累斯顿工业大学，德国
†信息论与机器学习教席，德累斯顿工业大学，德国
‡人机协同触觉互联网中心（CeTI），德国德累斯顿

{firstname.lastname}@tu-dresden.de

arXiv:2611.32661v1 [cs.NI] 24 Nov 2026

摘要—**本文提出记忆感知Q路由（MAQR），用于在多跳生物纳米物联网（IoBNT）网络中基于历史的分子链路上实现新鲜状态传递。重复传输会留下残留分子，使链路可靠性依赖于历史，并导致固定路由效率低下。本文基于空间马尔可夫模型对每条链路进行校准，并将其近似为三个非负衰减模态。由此得到的降维记忆状态和数据包比特组成可提供瞬时链路成功率估计。MAQR将量化记忆和可靠性描述符与数据包年龄、队列占用量及分子储备相结合，构成一种轻量级的部分可观测Q路由启发式方法。该方法离线训练，在部署期间仅需进行降维状态更新和Q表查找。在五组相互独立的训练对和20个共享评估种子上，与不考虑记忆的Q路由相比，MAQR在**0.30 **分组/时隙的传输速率下将平均信息年龄降低了**46.9% **，同时提高了分子效率。冻结策略扫描进一步表明，在弱记忆条件下两者性能相当，而随着信道记忆增强，MAQR的优势不断扩大。**

索引词—**信息年龄、生物纳米物联网、分子通信、分子信道记忆、Q路由、强化学习。**

I. 引言

THE 生物纳米物联网（IoBNT）构想了由生物纳米传感器、工程化细胞和生物-网络接口组成的网络，用于体内传感和执行控制 [1], [2]。分子通信（MC）为在如此微小的尺度上传输信息提供了一种具有生物相容性的方式 [3]。由于单个分子链路的覆盖范围有限，译码转发中继可以扩展通信覆盖范围，并支持将生理状态更新多跳传送至生物信息网关 [4], [5]。一个受生物学启发的例子是细菌纳米网络，其中可以利用运动性、趋化性和接合等自然特性，通过中间细菌群体传输信息 [6], [7]。

此类网络面临的一项关键挑战是，分子链路可靠性取决于近期的传输历史。扩散、平流和可逆接收会产生较长的信道拖尾，留下会影响后续观测的残留分子 [8]。因此，反复使用同一链路可能会降低其瞬时可靠性，使固定的最短路由或平均可靠性路由效率低下。由此产生的解码失败和重传会增加分子消耗，并延迟新鲜状态传递。这对于监测应用尤为重要，因为信息年龄（AoI）能够反映目的地可获得的最新更新的新鲜度 [9]。点对点分子信道已经表现出更新频率、码间干扰与信息年龄之间的直接权衡关系 [10]。

在更广泛的通信网络中，近期研究探讨了多通道多跳无线系统中的AoI优化、低功耗物联网基础设施中的新鲜度感知的下一跳选择，以及时变拓扑下基于学习的路由 [11], [12]。总体而言，这些研究突出了根据信息新鲜度和不断演化的网络状况调整转发决策的重要性。然而，现有的多跳分子通信研究主要集中于中继部署、跳数，以及基于平均链路质量或能量消耗的路由 [13]–[15]。机会式分子路由还利用距离和浓度梯度信息来选择中继节点 [16]。这些方法并未根据多个输出分子链路瞬时的传输诱导记忆状态，对路由决策进行显式自适应调整。将完整的物理信道状态直接暴露给路由层，也会给资源受限的物联网纳米生物节点带来过高的复杂性。因此，需要一种紧凑的分子记忆跨层表示，以支持历史感知路由。

为解决这一问题，我们提出了*memory-aware Q-routing* （MAQR），用于多跳 IoBNT 网络中的新鲜度感知的数据包转发。MAQR 采用残余分子记忆和链路可靠性的精简表示，

本研究得到德国研究联合会（DFG）的资助，属于德国卓越战略下的项目——EXC 2050/2——德累斯顿工业大学卓越集群“以人为中心的触觉互联网”（CeTI），项目编号为 390696704；同时得到德国联邦研究、技术与航天部（BMFTR）的资助，属于通信系统研究计划“Souver¨an. Digital. Vernetzt.”。联合项目 6G-life，项目编号 16KIS2413K；研究计划“联合项目：面向技术主权、韧性与共享繁荣的颠覆性通信范式——产业转化与创新技术演示平台建设——CommUnity”，项目编号 16KISS012K；以及项目“生物纳米物联网”（IoBNT），项目编号 16KIS1994。

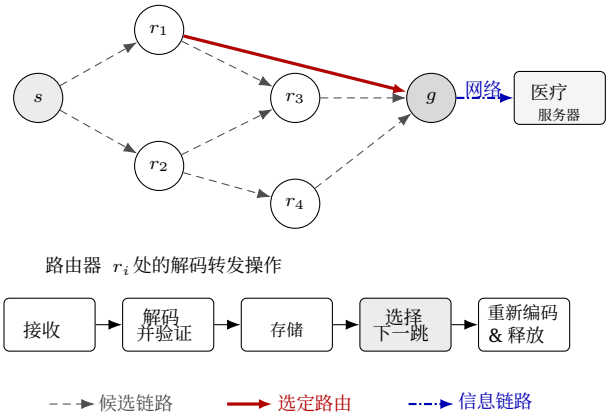


图 1. 多跳物联网纳米生物技术架构及解码转发操作。由于重复激活会留下残留分子，较短的路由可能暂时表现更差。链路参数表示为 $(d_{ij},\ v_{ij})$，其中距离单位为μ米，流速单位为 μ米/秒：$s\to r_1$：(18，6)；$s\to r_2$：(14，16)；$r_1\to r_3$：(15，6)；$r_2\to r_3$：(24，16)；$r_2\to r_4$：(16，16)；$r_1\to g$：(29，6)；$r_3\to g$：(14，16)；以及 $r_4\to g$：(18，16)。

并将其融入轻量级表格型Q路由框架中，同时通过离线训练避免在线探索。我们的主要贡献如下：

- 我们针对历史依赖型分子链路上的新鲜状态传递进行建模，采用一致的单比特响应、依赖于数据包的分子释放，以及组合条件下的数据包可靠性。
- 我们将每个空间马尔可夫信道简化为三个正衰减模态，并仅向路由层提供量化的残留记忆和可靠性描述符。
- 我们提出一种离线训练的表格型MAQR启发式算法，并将其与状态匹配的无记忆智能体以及动态非学习型记忆感知基线进行比较。

符号说明：斜体、粗体小写字母、粗体大写字母和花体字母分别表示标量、列向量、矩阵和集合。符号 $|\mathcal{A}|$、$(\cdot)^{\mathsf{T}}$和 $\mathbf{1}\{\cdot\}$ 分别表示基数、转置和指示函数；$[x]^+ \triangleq$ 最大值$\{x,0\}$。

II. 系统模型

A. 网络、数据包与资源模型

如图1所示，我们考虑一个由有向图 $\mathcal{G}=(\mathcal{V},\ \mathcal{E})$表示的多跳IoBNT，其中包含 $\mathcal{V}=\{s\}\cup\mathcal{V}_{\mathrm{R}}\cup\{g\}$，其中 s 是一个生物纳米传感节点，$\mathcal{V}_{\mathrm{R}}$是译码转发分子路由器的集合，g 是生物信息网关。有向边$(i,\ j)\in\mathcal{E}$ 表示一条可行的分子链路，而 $\mathcal{N}_i^+=\{j:(i,\ j)\in\mathcal{E}\}$表示节点 i的出邻居集合。每条链路$(i,\ j)$由其传播距离 d_{ij} 和流速 v_{ij}表征；它们与其余物理通道参数共同决定链路的传播特性和记忆特性。

我们考虑一个时隙网络，其中每个网络时隙的持续时间为 T_s。令 $A[t]\in\{0,1\}$ 表示感知节点 s 在时隙 t的更新到达指示变量，其中 $A[t]=1$表示在时隙开始时生成了一个新的状态更新，t表示其他情况。更新到达被建模为独立同分布的伯努利过程 [17]，即，

$$\Pr\{A[t]=1\}=\lambda,\qquad \Pr\{A[t]=0\}=1-\lambda, \tag{1}$$

其中，$\lambda=\mathbb{E}[A[t]]$ 表示每时隙分组的平均更新生成速率。感知节点采用最新更新替换策略：每当生成新的更新时，源节点处等待传输的任何未交付分组都会被新生成的分组替换。中间路由器采用容量有限的先进先出（FIFO）队列，容量为 $Q_{\max}$。

每个分组 p 表示为二进制序列

$$\boldsymbol{\beta}_p=\left[\beta_p[0],\ldots,\beta_p[B_{\mathrm{pkt}}-1]\right]^{\top},\qquad \beta_p[q]\in\{0,1\}, \tag{2}$$

其中，B_{pkt} 表示以比特计的分组长度。该分组在一个网络时隙内传输，由 B_{pkt} 个时长为 $T_b=T_s/B_{\mathrm{pkt}}$的等长比特间隔组成。比特 q 在其对应间隔开始时释放，并在固定的链路决策延迟 $\tau_{ij}^{\mathrm{dec}}<T_b$之后进行采样。分组比特通过二进制浓度移位键控（BCSK）在每条分子链路上传输，其中比特 1由释放 N_{tx}个信号分子表示，比特 0则不释放分子。为刻画由此产生的分子成本，我们定义

$$B_p^{(1)}=\sum_{q=0}^{B_{\mathrm{pkt}}-1}\beta_p[q] \tag{3}$$

为分组 p中比特 1 符号的数量。因此，一次传输尝试消耗数量为

$$M_p^{\mathrm{tx}}=B_p^{(1)}N_{\mathrm{tx}} \tag{4}$$

信号分子。每次分组传输尝试占用一个网络时隙。假设能够完美检测分组错误，例如通过理想的循环冗余校验（CRC）进行检测，其信令开销不作显式建模。

在网络层，分组转发通过链路激活决策进行控制。令 $\sigma_{ij}[t]\in\{0,1\}$ 表示节点 i 是否在时隙 t 激活链路 (i,j)。每个节点每个时隙至多激活一条出链路，即

$$\sum_{j\in\mathcal{N}_i^+}\sigma_{ij}[t]\le 1. \tag{5}$$

当 $\sigma_{ij}[t]=1$时，节点 i 尝试将其队首分组传输至节点 j。若该分组被成功解码且未检测到分组错误，则将其从发送端队列中移除并插入接收端队列；若 $j=g$，则该分组被视为已送达并离开网络。否则，该分组仍保留在节点 i，以便稍后重传。时隙 t 中的所有路由决策均基于该时隙开始时所观测到的队列状态。成功到达的分组在所有时隙起始决策处理完毕后才被追加，转发至中间路由器的分组仅在下一时隙才具备传输资格。若同时到达的分组超过剩余缓冲区容量，则按生成时间降序、再按序列标识符排序；超出的分组将被丢弃。

除链路激活外，分组传输还受各发送节点可用信号分子储备的约束。令 $E_i[t]$ 表示节点 i 在时隙 t补充前的储备，$E_{\max}$ 表示预留容量，E_{rec}表示每时隙补充的分子数。可用于传输的储备为

$$\widetilde{E}_i[t] = \min\left\{E_{\max}, E_i[t] + E_{\mathrm{rec}}\right\}. \tag{6}$$

~只有在 $E_i[t] \geq M_p^{\mathrm{tx}}$时，分组 p 才能被传输。根据传输决策，储备量的变化为

$$E_i[t+1] = \widetilde{E}_i[t] - M_p^{\mathrm{tx}} \sum_{j \in \mathcal{N}_i^+} \sigma_{ij}[t], \tag{7}$$

由于式(5)中的求和项只能为零或一，因此无论解码结果如何，每次传输尝试都会消耗 M_p^{tx}个分子，而空闲节点则保留其补充后的储备量。

在全文中，假设逐跳确认无错误，且传输持续时间可忽略不计。通过使用可区分的分子类型或互不重叠的传输时序，将可能相互干扰的分子链路进行正交化。因此，忽略链路间干扰，每条有向链路的残余状态仅取决于其自身的传输历史。

B. 物理链路和简化记忆模型

首先，使用 [18]–[20]中的空间瞬态马尔可夫模型，对每条有向分子链路（i，j）$\in \mathcal{E}$ 进行表征。在物理时间步 k，瞬态分子状态演化为

$$\boldsymbol{x}_{ij}[k+1] = \boldsymbol{P}_{ij}\left(\boldsymbol{x}_{ij}[k] + \boldsymbol{b}_{ij} u_{ij}[k]\right), \tag{8}$$

其中，$\boldsymbol{x}_{ij}[k]$ 包含瞬态空间状态和接收器状态中的期望分子数，$\boldsymbol{P}_{ij}$ 为瞬态转移矩阵，$\boldsymbol{b}_{ij}$ 指定注入状态，$u_{ij}[k]$表示分子输入。转移矩阵 $\boldsymbol{P}_{ij}$ 由扩散系数、链路几何结构、固定流速、可逆结合动力学和下游流出共同确定。该完整空间模型仅用于离线链路表征，不由路由算法进行传播。

我们定义 $\boldsymbol{o}_{ij}$ 为选择接收端观测状态的向量，$\boldsymbol{A}_{ij}^{\mathrm{dec}}$ 和 $\boldsymbol{A}_{ij}^{\mathrm{slot}}$ 分别为在 τ_{ij}^{dec} 和一个网络时隙内的传播矩阵。在这一统一的物理尺度上，采样得到的 N_{tx} 个分子发生比特1释放时的响应为

$$\bar{h}_{ij}[n] = N_{\mathrm{tx}} \boldsymbol{o}_{ij}^{\mathsf{T}} \boldsymbol{A}_{ij}^{\mathrm{dec}} (\boldsymbol{A}_{ij}^{\mathrm{slot}})^n \boldsymbol{b}_{ij}, \qquad n \geq 0. \tag{9}$$

¯样本 $h_{ij}[0]$ 是一个比特-¯1 所产生的期望贡献均值，而 $\{h_{ij}[n]\}_{n\geq 1}$ 表征其网络时隙后游标记忆。 [18]中的转移矩阵特征分解表明，分子信道响应可采用模态表示。因此，我们使用 R 个非负衰减模态来近似与公式（8）相关的后游标尾部，

$$\bar{h}_{ij}[n+1] \approx \sum_{r=1}^{R} w_{ij,r} p_{ij,r}^{n}, \qquad n \geq 0, \tag{10}$$

其中$w_{ij,\,r} \geq 0$ 和$0 \leq p_{ij,\,r} < 1$ 分别表示模态权重和衰减因子。这些参数通过在$n = 0$、.....、$N_h - 1$上进行有界加权非线性最小二乘法获得，并使用 $1/(n+1)$作为平方残差权重。在同一校准窗口上，使用归一化均方根误差（NRMSE）和相对 ℓ_2 误差评估每条链路的近似精度。对于 NRMSE，将均方根拟合误差除以最大的物理尾部样本进行归一化。

在网络时隙尺度上，拟合得到的衰减因子和模态权重分别收集于 $\boldsymbol{\Lambda}_{ij} = \mathrm{diag}(p_{ij,\,1}, \ldots, p_{ij,\,R})$ 和 $\boldsymbol{w}_{ij} = [w_{ij,1}, \ldots, w_{ij,R}]^{\top}$ 中。随后，降阶后的残余状态演化为

$$\boldsymbol{z}_{ij}[t+1] = \boldsymbol{\Lambda}_{ij} \boldsymbol{z}_{ij}[t] + \boldsymbol{w}_{ij} B_p^{(1)} \sigma_{ij}[t]. \tag{11}$$

因此，每次尝试分组传输都会叠加由其 $B_p^{(1)}$ 次比特1释放产生的校准尾部，而不受解码结果影响；空闲链路则仅发生模态衰减。

时隙 t 中路由决策可用的残余分子贡献可直接由降阶模态状态得出：

$$\rho_{ij}[t] \triangleq \mathbf{1}^{\top} \boldsymbol{z}_{ij}[t], \tag{12}$$

重复的链路激活会增加残余贡献，而空闲时隙则使模态状态衰减。为量化每条物理链路固有的记忆强度，我们定义离线记忆指数

$$\chi_{ij} \triangleq \frac{\sum_{n=1}^{N_h} \bar{h}_{ij}[n]}{\sum_{n=0}^{N_h} \bar{h}_{ij}[n]}, \tag{13}$$

其中，N_h 表示链路校准期间使用的脉冲响应时域。较大的 χ_{ij} 表示有更大比例的分子响应会持续到第一个网络时隙观测之后。对网络中的链路级指数进行汇总可得

$$\overline{\chi} = \frac{1}{|\mathcal{E}|} \sum_{(i,j)\in\mathcal{E}} \chi_{ij}, \tag{14}$$

较大的χ 对应更强的整体分子记忆效应。

C. 记忆相关可靠性与跨层状态

给定时隙 t中路由决策前的残余分子贡献 $\rho_{ij}[t]$，比特 0 和比特 1 假设下的接收端观测建模为

$$Y_{ij}^{(0)}[t] \sim \mathrm{Poisson}\left(\lambda_{ij}^{\mathrm{bg}} + \rho_{ij}[t]\right), \tag{15a}$$

$$Y_{ij}^{(1)}[t] \sim \mathrm{Poisson}\left(\lambda_{ij}^{\mathrm{bg}} + \rho_{ij}[t] + s_{ij}\right), \tag{15b}$$

其中，$\lambda_{ij}^{\mathrm{bg}}$ 表示背景分子计数，

$$s_{ij} = \bar{h}_{ij}[0]$$

表示经校准的单比特期望贡献。每个整数阈值 η_{ij} 均在路由评估前通过最小化孤立链路比特错误确定，并由所有策略共享。

对于当前的分组前残差，定义虚警概率和漏检概率

$$P_{ij}^{\mathrm{FA}}[t] = \Pr\{Y_{ij}^{(0)}[t] \geq \eta_{ij}\}, \tag{16a}$$

$$P_{ij}^{\mathrm{M}}[t] = \Pr\{Y_{ij}^{(1)}[t] < \eta_{ij}\}. \tag{16b}$$

在给定实际分组组成的条件下，独立比特分组成功近似变为

$$P_{ij,p}^{\mathrm{s}}[t] = \left(1 - P_{ij}^{\mathrm{FA}}[t]\right)^{B_{\mathrm{pkt}} - B_p^{(1)}} \left(1 - P_{ij}^{\mathrm{M}}[t]\right)^{B_p^{(1)}}. \tag{17}$$

该网络时隙抽象对所有比特判决使用相同的分组前残差 $\rho_{ij}[t]$，并在式（11）中汇总分组的比特 1 后游标；它不会显式解析分组内状态演化。因此，我们将其网络时隙残差和分组成功轨迹与相应的完整空间马尔可夫响应进行对比验证。因此，瞬时可靠性取决于近期链路使用情况以及队首分组的组成。

图 2 展示了近期对代表性慢清除链路 $s \to r_1$ 的使用如何影响其可靠性。在连续传输期间，残余分子贡献不断累积，并逐步降低受分组组成影响的分组成功概率，尽管分组组成和静态链路参数保持不变。传输停止后，模态衰减逐渐清除残差记忆，使分组成功概率恢复至接近链路清除后的数值。简化模型的标记点与相应的完整空间曲线几乎重合。对于链路 $s \to r_1$，残差轨迹的相对 ℓ_2 误差为 2.58×10^{-4}，而分组成功概率的最大绝对误差为 1.45×10^{-2}。在全部八条有向链路上，$R = 3$ 尾部近似的归一化均方根误差均不超过 3.07×10^{-4}。这些结果表明，简化状态保留了相关的累积与清除行为。因此，分子链路可靠性不仅取决于静态物理属性，还取决于近期传输历史。

为将这种历史依赖性提供给路由层，同时无需传播完整的物理马尔可夫状态，每条出链路都会提供量化跨层描述符

$$c_{ij}^{\rho}[t] = \mathcal{Q}_{\rho}\left(\rho_{ij}[t]\right), \qquad c_{ij,p}^{\mathrm{s}}[t] = \mathcal{Q}_{\mathrm{s}}\left(P_{ij,p}^{\mathrm{s}}[t]\right). \tag{18}$$

其中 $\mathcal{Q}_{\rho}(\cdot)$ 和 $\mathcal{Q}_{\mathrm{s}}(\cdot)$ 分别将残差记忆和分组成功概率映射为三类。因此，$c_{ij}^{\rho}[t]$，$c_{ij,\,p}^{\mathrm{s}}[t] \in \{0,1,2\}$。每个链路描述符的两个阈值均为经验三分位数，由独立的标称负载校准轨迹得出，并在此前固定 Q 训练与评估。尽管 $c_{ij,\,p}^{\mathrm{s}}[t]$ 部分源自 $\rho_{ij}[t]$，但它还捕获了分组组成以及链路特定的信号和阈值参数。这些描述符共同提供了紧凑的物理到网络接口，而无需暴露空间马尔可夫状态。

III. 新鲜度感知路由问题表述

本节从分组新鲜度、传输截止期限、局部观测和路由动作等方面，阐述新鲜度感知路由问题。对于分组，p t_p^{gen}表示

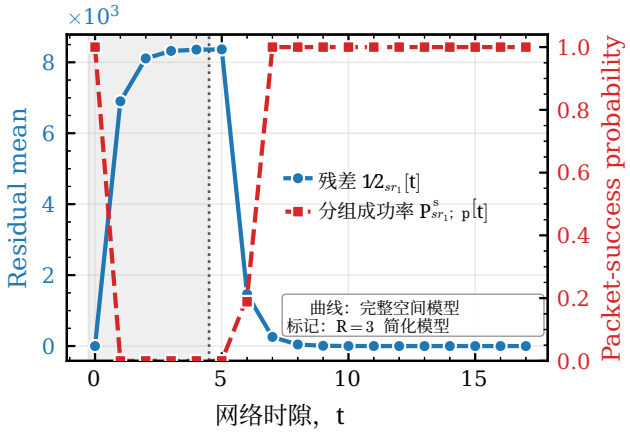


图 2. 链路 $s \to r_1$ 上的残余记忆累积与空闲清除。曲线根据公式 (8) 中的完整空间模型获得，而标记根据公式 (11) 中的 $R = 3$ 简化模型获得。从清除状态开始，每个时隙传输一个包含 $B_p^{(1)} = 8$ 个比特 1 符号的分组，持续 $t = 0, \ldots\ldots, 4$ 个时隙，随后进入空闲时隙。

其生成时隙，t_p^{del}表示其在网关成功送达的时隙；对于尚未送达的分组，$t_p^{\mathrm{del}} = \infty$。由于传输结果在网络时隙结束时才确定，因此在时隙 t 开始时，网关处最新更新的生成时间为

$$U[t] = \max\left(\left\{t_p^{\mathrm{gen}} : t_p^{\mathrm{del}} < t\right\} \cup \{U_0\}\right), \tag{19}$$

其中，$U_0 \leq 0$ 表示网关处最初可用的更新的生成时间。网关处可用信息的新鲜度通过信息年龄 AoI 衡量，

$$\Delta[t] = t - U[t], \tag{20}$$

它会随着自最新一次更新传输以来经过的时间而增加。虽然 $\Delta[t]$ 刻画了目的端新鲜度，但路由决策必须针对仍在网络中的分组。因此，我们使用未传输分组在时隙 t开始时的年龄 p，

$$\tau_p[t] \triangleq t - t_p^{\mathrm{gen}}, \tag{21}$$

作为衡量其传输紧迫性的局部指标。分组传输受端到端截止期限 τ_{dl}的约束，该期限以网络时隙计。若分组在时隙 t_p^{del}传输，则视为按时传输，当

$$\tau_p[t_p^{\mathrm{del}}] + 1 = t_p^{\mathrm{del}} - t_p^{\mathrm{gen}} + 1 \leq \tau_{\mathrm{dl}}. \tag{22}$$

因此，每当 $\tau_p[t] \geq \tau_{\mathrm{dl}}$时，未传输的分组都会在时隙 t 开始时从网络中移除。

为区分真正刷新网关的传输与乱序过时传输，令 $\mathcal{D}_g[t]$ 表示在时隙 t结束时传输的分组集合。若该集合非空，则令 $p_t^{\star}$为其中生成时间最晚的分组（若生成时间相同，则以序列标识符判定）。我们定义

$$\begin{aligned} I_p^{\mathrm{info}}[t] &= \mathbf{1}\left\{p = p_t^{\star},\ t_p^{\mathrm{gen}} > U[t]\right\}, \\ G_p^{\mathrm{AoI}}[t] &= I_p^{\mathrm{info}}[t]\left(t_p^{\mathrm{gen}} - U[t]\right), \end{aligned} \tag{23}$$

其中，当 $\mathcal{D}_g[t] = \varnothing$ 时，这两个量均为零。因此，只有最新鲜的同时交付才能推进网关时间戳 $U[t+1] = \max\{U[t], \max_{p\in\mathcal{D}_g[t]} t_p^{\mathrm{gen}}\}$，而 $G_p^{\mathrm{AoI}}[t]$ 是

相较于未收到更新，由此产生的结果是减少一个时隙。此约定可防止过时或被支配的同时数据包获得新鲜度收益。

为评估在 T 个时隙的测量区间内所得的路由性能，我们首先计算网关平均信息年龄：

$$\overline{\Delta} = \frac{1}{T}\sum_{t=1}^{T}\Delta[t], \tag{24}$$

该指标反映网关可用状态信息的长期新鲜度。及时性进一步通过在 τ_{dl} 内到达网关并推进其时间戳的生成数据包所占比例进行量化。令 $\mathcal{P}_{\mathrm{gen}}$ 表示测量区间内生成的分组集合，$\mathcal{P}_{\mathrm{on}}^{\mathrm{info}} \subseteq \mathcal{P}_{\mathrm{gen}}$ 表示根据公式（23）交付及时且具有信息性的分组子集。信息性截止期限交付比率为

$$P_{\mathrm{D}}^{\mathrm{info}} = \frac{|\mathcal{P}_{\mathrm{on}}^{\mathrm{info}}|}{|\mathcal{P}_{\mathrm{gen}}|}, \tag{25}$$

该指标补充了 Δ，且不会将乱序过时到达视为有用的状态更新。

由于新鲜且及时的交付可能需要不同次数的传输尝试，我们还测量相应的分子消耗。令 $\mathcal{A}_T$ 表示同一测量区间内的传输尝试集合，$p(\nu)$表示与尝试 $\nu \in \mathcal{A}_T$相关联的分组。每次信息性及时交付所消耗的平均分子数为

$$\overline{M}_{\mathrm{on}}^{\mathrm{info}} = \frac{\sum_{\nu\in\mathcal{A}_T} M_{p(\nu)}^{\mathrm{tx}}}{|\mathcal{P}_{\mathrm{on}}^{\mathrm{info}}|}, \tag{26}$$

该指标涵盖所有尝试的传输，包括未成功的传输，以及随后过期或被丢弃的分组的传输。

接下来，我们说明可用于路由决策的局部信息。令 $L_i[t]$ 表示节点 i的队列占用量。对于队首分组 p，节点 i 构造量化局部描述符 $c_p^A[t] = \mathcal{Q}_A(\tau_p[t])$、$c_i^q[t] = \mathcal{Q}_q(L_i[t])$ 和 $c_i^E[t] = \mathcal{Q}_E(\widetilde{E}_i[t]/E_{\max})$，其中 $\mathcal{Q}_A(\cdot)$、$\mathcal{Q}_q(\cdot)$ 和$\mathcal{Q}_E(\cdot)$分别将数据包年龄、队列占用量和归一化分子储备映射到离散路由状态类别。将这些局部描述符与式（18）中的链路级记忆和可靠性描述符相结合，可得到节点 i处的 MAQR观测：

$$\boldsymbol{o}_{i,p}[t] = \left(i, c_p^A[t], c_i^q[t], c_i^E[t], \left\{c_{ij}^{\rho}[t], c_{ij,p}^{\mathrm{s}}[t]\right\}_{j\in\mathcal{N}_i^+}\right), \tag{27}$$

其中，节点标识符 i用于区分共享路由策略中的节点特定状态。由此得到的观测共同反映了数据包紧迫性、局部拥塞、分子可用性，以及每条候选出链路的记忆相关可靠性。

每当节点 i 持有分组且分子储备足以进行传输时，该观测被映射为一个下一跳动作，

$$a_i[t] = \pi\left(\boldsymbol{o}_{i,p}[t]\right) \in \mathcal{N}_i^+, \tag{28}$$

从而据此激活所选的出链路

$$\sigma_{ij}[t] = \mathbf{1}\{a_i[t] = j\}. \tag{29}$$

如果队列为空或分子储备不足，则不会激活任何出链路。任何传输尝试都会消耗分子，并更新所选链路的残余记忆状态，无论解码是否成功。

IV. 记忆感知Q路由

A. 离线Q学习

经典Q路由使用本地Q值选择前往目的地 [21]的下一跳。我们保留这种轻量级表格结构，同时利用式（27）中的分子记忆描述符增强路由观测。由于节点标识符包含在$\boldsymbol{o}_{i,p}[t]$中，共享Q函数可以表示特定于节点的路由决策。MAQR使用精简的分子链路模型进行离线训练，从而避免在部署期间进行分子成本高昂的探索和在线Q值更新。

先进先出转发要求延续状态属于同一个分组。对于在时隙 t 的节点 i 处针对分组 p 执行的动作，设 $d_{i,p}[t] \geq 1$ 表示直到 p 成为队首并具备进行路由决策资格所需的时隙数，并设 $(i_p^+, \boldsymbol{o}_{i_p^+,p}^+)$ 表示该节点及其观测。传输失败通常令 $d_{i,p}[t] = 1$ 和 $i_p^+ = i$。成功转发至非终端下一跳 j后，只有当 p 在 j处成为队首时才进行引导更新，而不是使用 $t+1$处另一个分组的状态。网关交付、过期和缓冲区丢弃均为终止状态。事件驱动更新为

$$\begin{aligned} Q(\boldsymbol{o}_{i,p}[t], j) \leftarrow Q(\boldsymbol{o}_{i,p}[t], j) + \alpha_t \Big[& \mathcal{R}_{i,p}[t, d_{i,p}] \\ & + \gamma^{d_{i,p}} J_{i,p}[t] \max_{j'\in\mathcal{N}_{i_p^+}^+} Q\left(\boldsymbol{o}_{i_p^+,p}^+, j'\right) - Q(\boldsymbol{o}_{i,p}[t], j) \Big], \end{aligned} \tag{30}$$

其中 $\alpha_t \in (0, 1]$，$\gamma \in [0, 1)$，且 $J_{i,p}[t]$ 仅当存在针对 p 的后续决策时才为 1。转移回报 $\mathcal{R}_{i,p}[t, d_{i,p}]$ 包含下述传输奖励，以及在下一次决策之前产生的任何折扣过期或丢弃惩罚。本次更新不使用其他分组的奖励或观测。

仅用于成本归一化，定义记账参考值 $M_{\mathrm{ref}}^{\mathrm{cost}} = (B_{\mathrm{pkt}}/2)N_{\mathrm{tx}}$。它不用于对公式（9）中的物理响应进行缩放。然后

$$\widetilde{M}_p^{\mathrm{tx}} \triangleq \frac{M_p^{\mathrm{tx}}}{M_{\mathrm{ref}}^{\mathrm{cost}}} = \frac{2B_p^{(1)}}{B_{\mathrm{pkt}}}. \tag{31}$$

对于节点 i 在时隙 t 中传输的分组 p，令 $F_i[t]$ 表示传输失败，$X_i[t]$ 表示因过期或缓冲区丢弃导致的数据包终止。即时的数据包特定奖励为

$$\begin{aligned} R_{i,p}[t] = c_g \frac{G_p^{\mathrm{AoI}}[t]}{\tau_{\mathrm{dl}}} - c_h - c_M \widetilde{M}_p^{\mathrm{tx}} \\ - c_q[L_i[t] - 1]^+ - c_f F_i[t] - c_e X_i[t]. \end{aligned} \tag{32}$$

第一项仅奖励能够推进$U[t]$ 的信息性送达，并根据由此带来的信息年龄降低进行缩放，而

其余各项则考虑跳数使用量、归一化分子开销、排队数据包、传输失败和数据包终止。观测中保留数据包年龄，使紧迫性在送达前影响转发。由于决策依赖局部信息，而非完整的网络信息，式（30）被解释为一种用于去中心化运行的轻量级、部分可观测、事件驱动型Q路由启发式方法，其有效性通过配对实证评估进行检验。

在离线训练期间，ε-贪心策略会平衡探索与利用，其中 ε 和 α_t 均会在训练回合数中线性下降。流量负载从预定义的训练集抽取，初始链路记忆状态则随机化，以使学习器接触清空和已加载记忆的两种状况。在部署期间，探索被禁用，下一跳选择为

$$a_i^\star[t] = \arg\max_{j \in \mathcal{N}_i^+} Q\left(\boldsymbol{o}_{i,p}[t], j\right). \tag{33}$$

对于离线训练期间未遇到的观测，策略将回退到下文所述的瞬时记忆感知路由（IMR）规则。

B. 基线路由策略

为进行比较，除MAQR外，我们还考虑了三种参考路由策略。最短跳数路由（*SHR*）选择使剩余跳数最小的固定下一跳，仅使用清空链路可靠性来打破平局。非学习型瞬时记忆感知路由（*IMR*）基线根据与MAQR可获得的相同成分条件链路可靠性，重新计算动态负对数路径代价：

$$\begin{aligned} C_{i,p}^{\mathrm{IMR}}[t] = \min_{j \in \mathcal{N}_i^+} \big[& -\log P_{ij,p}^{\mathrm{s}}[t] + \kappa_h \\ & + \kappa_M \widetilde{M}_p^{\mathrm{tx}} + C_{j,p}^{\mathrm{IMR}}[t] \big], \\ C_{g,p}^{\mathrm{IMR}}[t] = 0, & \end{aligned} \tag{34}$$

其中，κ_h 和 κ_M 分别对额外跳数及其分子代价施加惩罚。DAG代价按逆拓扑顺序计算。IMR在每次失败后重新评估，并且能够感知记忆状态，但不使用数据包年龄、队列占用量、分子储备或学习。因此，它用于检验Q路由是否能够在简单的瞬时可靠性启发式方法之外带来额外价值。

第三种参考策略，即记忆盲*Q*路由（*MBQR*），采用与MAQR相同的事件驱动更新、奖励、训练种子和局部非链路观测，但将瞬时链路描述符替换为在独立的标称负载校准轨迹上测得的每条出链路的量化平均可靠性。相比之下，MAQR同时观测 $c_{ij}^{\rho}[t]$和$c_{ij,\ p}^{\mathrm{s}}[t]$。因此，配对的MAQR–MBQR比较仅改变了对瞬时链路记忆信息的访问方式。

C. 在线实现复杂度

在部署期间，MAQR 的计算成本主要由降维状态跟踪和 Q表查找构成。对于出度为 $d_i = |\mathcal{N}_i^+|$ 的节点 i，更新所有候选出链路的 R 个模态状态需要 $O(Rd_i)$ 次 Q 值比较以完成下一跳选择。由于 $R = 3$ 是固定的，在线计算成本与局部节点度呈线性增长，并且与式（8）中底层空间马尔可夫模型的维度无关。

表I 主要物理、流量、学习和评估参数。

类别	参数	值
通道	$D, \Delta x, \Delta t$	150 μm^2/s　1 μm 2.5 毫秒
	$k_{\mathrm{on}}, c_p, k_{\mathrm{off}}$	6×10^8 M^{-1}s^{-1}，10 nM，3 s^{-1}
	连接距离 d_{ij}	14–29 μ米
	连接流速 v_{ij}	6，16 μ米/秒
	吸收缓冲区	5 μm
	网络时隙持续时间 T_s	2 s
	比特间隔 T_b	0.125 秒
	背景均值 λ^{bg}	4060
	阈值 η_{ij} 链路级孤立比特最优值	
	模态阶数 / 拟合时域（R，N_h）	（3，15）
流量	数据包长度 B_{pkt}	16 比特
	每比特分子数-1 N_{tx}	10^5
	端到端截止期限 τ_{dl}	10 个时隙
	路由器队列容量 $Q_{\max}$	4 个分组
	预留容量 $E_{\max}$	8×10^6 分子
	恢复 E_{rec}	6×10^5 分子/时隙
	评估速率 λ	$\{0.05, 0.10, \ldots, 0.30\}$ 分组/时隙
学习	训练速率 λ_{train}	$\{0.08, 0.14, 0.20, 0.26, 0.30\}$ 分组/时隙
	$\mathcal{Q}_A/\mathcal{Q}_q$ 阈值	(3，7) 时隙/ (1，3) 分组
	$\mathcal{Q}_E$ 阈值	(0.25, 0.65)
	$\mathcal{Q}_\rho, \mathcal{Q}_s$ 阈值	链路级名义三分位数
	折扣因子 γ	0.95
	学习率 α	$0.50 \to 0.05$
	探索率 ε	$1.00 \to 0.02$
	$(c_q, c_h, c_M, c_\rho, c_f, c_e)$	(18, 1, 0.25,0.20,3, 6)
	IMR（κ_h，κ_M）	(1, 0.25)
	训练回合数 / 长度	5000 回合/ 220 时隙
	独立训练随机种子	5
评估预热 / 测量时长		$10^4/10^5$ 时隙
	成对评估种子	20

量化观测空间的大小取决于分组年龄、队列、储备、剩余内存和可靠性类别的数量，分别记为 C_A、C_q、C_E、C_ρ和 C_{s}。对于节点 i，可能的MAQR观测数量上限为

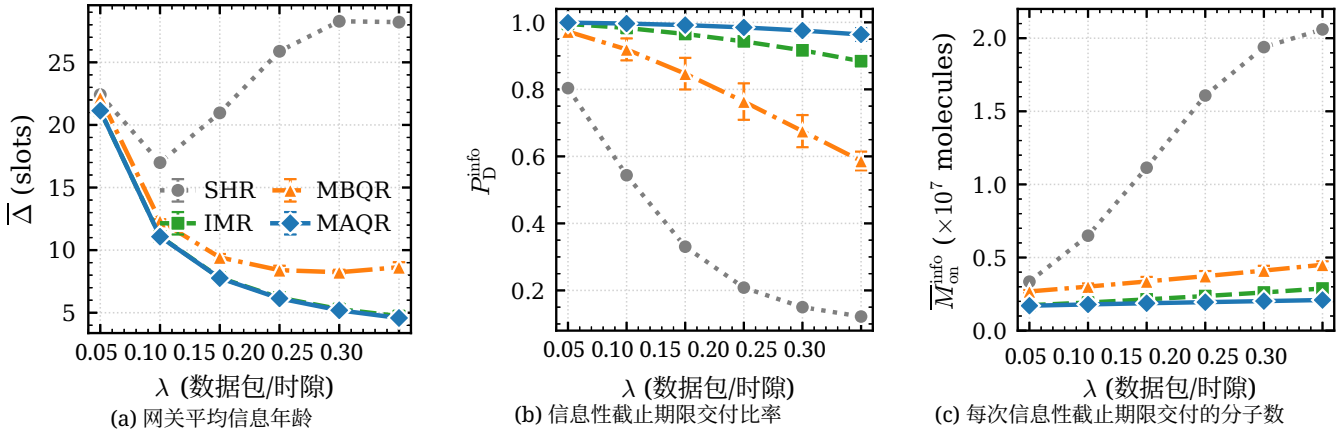
$$S_i^{\max} = C_A C_q C_E (C_\rho C_{\mathrm{s}})^{d_i}, \tag{35}$$

其中，$(C_\rho C_{\mathrm{s}})^{d_i}$ 表示所有候选出链路的联合内存和可靠性类别。由于节点标识符包含在$\boldsymbol{o}_{i\text{中},\ p}[t]$，不同路由节点的观测在共享Q函数中仍然可以区分。实际上，只需存储离线训练期间遇到的状态-动作对。在五组训练组合中，稀疏MBQR表和MAQR表分别包含95–108个和353–394个已访问条目，均值分别为100。8和378。6。因此，在线部署只需为每条出链路使用一个稀疏Q表和 R维模态状态，无需传播完整的空间马尔可夫状态，也无需执行神经网络推理。

V. 数值结果

A. 仿真设置

我们考虑图 1 所示的六节点拓扑，其中连接距离范围为 14 至 29 μm。链路 $(s,\ r_1)$ 和 $(r_1,\ g)$ 的流速为 6 μm/s，而其余有向链路的流速为 16 μm/s，流动方向均朝向接收



(a) 网关平均信息年龄 (b) 信息性截止期限交付比率 (c) 每次信息性截止期限交付的分子数

图 3. SHR、IMR、MBQR 和 MAQR 在不同更新生成速率 λ 下的路由性能。误差棒表示分层 95% 自助法置信区间。

节点。这在较短但内存敏感的路由 $s \to r_1 \to g$ 与较长但清空速度更快的替代路由之间形成了路由权衡。路由采用公式 (11) 中的简化模型，而公式 (8) 中的完整空间模型用于离线校准以及图 2 中的验证。除非另有说明，我们的自定义 Python 仿真器均采用表 I 中的参数。

离线训练期间，每个回合都会抽取 $\lambda_{\text{train}} \in \{0.08, 0.14, 0.20, 0.26, 0.30\}$ 和随机化的初始残余状态。五组相互独立的 MAQR–MBQR 训练种子对共享场景随机性，但使用相互独立的探索流。描述符阈值和 MBQR 平均可靠性根据单独的标称条件校准轨迹获得。所有流量负载和物理内存设置下，生成的策略均保持冻结。每组训练策略对使用 20 个共同评估种子，并针对分组到达、组成和链路结果使用共同随机数，在 10^4 个时隙的预热后评估 10^5 个测量时隙。误差棒表示跨训练策略对和评估种子的分层 95% 自助法置信区间；统计比较则基于配对策略差异。

B. 流量负载下的路由性能

图 3 使用式 (24) 至 (26) 中定义的指标，对比了更新生成率为 $\lambda = 0.05$–0.30 分组/时隙时的 SHR、IMR、MBQR 和 MAQR。图 3(a) 展示了平均信息年龄。随着 λ增加，MAQR 和 IMR 得益于更频繁的状态更新，信息年龄持续下降。相比之下，MBQR 的性能提升在高流量下趋于饱和，而 SHR 在超过 $\lambda = 0.10$ 后性能恶化，表明其固定的短路径无法适应不断增加的流量和残余记忆累积。在 $\lambda = 0.30$ 时，MAQR 的平均信息年龄为 4.58 个时隙，相比之下，MBQR 为 8.63 个时隙，IMR 为 4.73 个时隙。与 MBQR 相比，这降低了 46.9%；同时，MAQR 的信息年龄仍比瞬时可靠性基线低 0.149 个时隙。

图 3(b) 展示了信息性截止期限交付比率。MAQR 在整个负载范围内保持 $P_{\text{D}}^{\text{info}} \geq 0.964$，这表明即使流量不断增加，大多数生成的更新仍能及时交付且包含信息。相比之下，MBQR 和 SHR 的比率则随着负载增加而大幅下降。在 $\lambda = 0.30$ 时，MAQR 达到 $P_{\text{D}}^{\text{info}} = 0.964$，而 MBQR 和 IMR 分别达到 0.586 和 0.884。与 MBQR 之间不断扩大的差距表明，随着传输频率不断提高，瞬时内存描述符变得愈发重要。

图3(c)评估了每次信息性截止期限交付的分子消耗。MAQR在所有评估负载下的分子消耗均最低，表明其信息年龄改善并非通过不加区分地增加分子消耗实现。在 $\lambda = 0.30$ 时，MAQR每次信息性截止期限交付需要 2.09×10^6 个分子，而MBQR和IMR分别需要 4.50×10^6和 2.88×10^6 个分子。因此，相较于MBQR，MAQR的分子消耗降低了53.5%；相较于IMR，则减少了0.789×10^6 个分子。

在所有评估负载和指标上，成对的95%自助法置信区间均表明MAQR优于MBQR和IMR。未见观测回退比例低于学习策略传输尝试的 2.8×10^{-4} 。在 $\lambda = 0.20$时，MAQR在低记忆类别和高记忆类别中选择短源链路的条件概率分别为1.00和0.00，证实了依赖记忆的路由切换。

C. 冻结策略性能与物理记忆

为以可控方式改变物理信道记忆，我们引入记忆尺度 ζ，并将分子解离速率和链路流速缩放为

$$k_{\text{off}}(\zeta) = \frac{k_{\text{off}}}{\zeta}, \qquad v_{ij}(\zeta) = \frac{v_{ij}}{\sqrt{\zeta}}. \tag{36}$$

之所以改变可逆解离和平流，是因为它们决定了底层马尔可夫信道 [19] 中的分子滞留和清除。为将尾部持续性与清除后链路强度分离，我们在拟合约化模态之前对每个响应进行归一化：

$$\bar{h}_{ij}^{\text{sweep}}[n;\zeta] = \bar{h}_{ij}[n;\zeta]\frac{\bar{h}_{ij}[0;1]}{\bar{h}_{ij}[0;\zeta]}. \tag{37}$$

此归一化将单比特采样固定在其标称值，同时保留依赖于 ζ 的尾部形状，并且

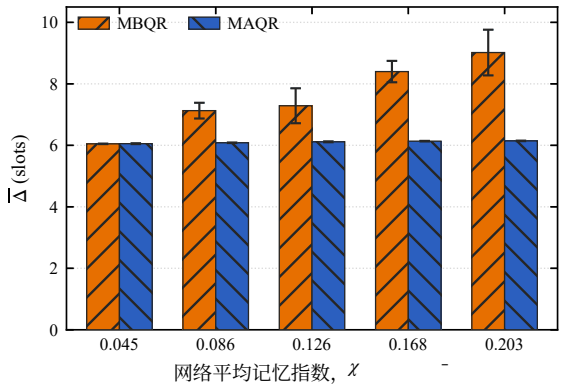


图4。在 $\lambda = 0.20$ 分组/时隙下，冻结 MBQR 和 MAQR 的网关平均信息年龄与网络平均记忆指数 $\bar{\chi}$ 的关系。误差棒表示分层 95% 自助法置信区间。

仅应用于受控记忆扫描。我们在 $\lambda = 0.20$ 分组/时隙下评估 $\zeta \in \{0.20, 0.45, 0.70, 1.00, 1.25\}$，并根据式 (13) 重新计算 $\bar{\chi}$。

图4显示，在物理记忆较弱时，MBQR 和 MAQR 的性能相当，但随着记忆持久性增加，两者逐渐分化。在最弱记忆点 $\bar{\chi} = 0.045$ 处，MBQR 和 MAQR 分别达到 6.049 和 6.054 时隙，其配对 95% 置信区间包含零。其余四个记忆点的配对区间均排除零，且方向 favoring MAQR。在最强记忆点 $\bar{\chi} = 0.203$ 处，MBQR 和 MAQR 分别达到 9.018 和 6.145 时隙，对应于降低 2.874 个时隙的信息年龄，即 31.9% 的信息年龄降低。

在记忆性最弱的情况下，MBQR 的未见观测比例达到 91.1%，在标称条件下低于 0.001%，而在记忆性最强的情况下升至 18.2%；相比之下，MAQR 的比例始终低于 0.12%。由于两个智能体对未见观测均采用相同的动态IMR规则，因此该图反映的是完整的冻结策略部署性能，包括状态覆盖和回退效应，而非仅基于描述符的因果比较。

VI. 结论

本文将MAQR构建为一种轻量级、部分可观测的Q路由启发式算法，用于基于历史依赖型分子链路的新鲜状态传递。在单比特校准和组分条件可靠性的支持下，MAQR将物理链路记忆与路由决策联系起来；同时，事件驱动先进先出更新和信息年龄降低奖励使学习过程与信息新鲜度保持一致。在评估的最高负载0.30分组/时隙下，与MBQR相比，MAQR将平均信息年龄降低了46.9%，同时提高了有效截止期限传递比率和分子效率。与IMR相比，MAQR也表现出幅度较小但稳定的改进。冻结策略记忆扫描进一步表明，在记忆较弱时，MAQR仍能保持相当的性能；随着物理记忆增强，MAQR的优势也不断扩大。这些结果证明了简化链路记忆状态对于自适应分子路由的价值。

未来的工作将使用完整的空间模型和实验校准的分子信道验证冻结策略，并量化其对模型阶数、物理模型失配、奖励设计和描述符选择的稳健性。此外，还将针对更大规模的多源IoBNT网络，研究可扩展的函数逼近、多智能体学习，以及对时变流量和信道条件的在线适应。

参考文献

[1] I. F. Akyildiz 等，“生物纳米物联网”，IEEE Commun. Mag.，第 53 卷，第 3 期，第 32–40 页，2015 年。[2] C. Lee 等，“血管中的生物纳米物联网：系统设计与原型”，J. Commun. Netw.，第 25 卷，2023 年，第 231 页，第 2 期，第 222 页。[3] M. Kuscu 等，“分子通信的发射机和接收机架构：关于调制、编码和检测技术物理设计的综述”，Proc. IEEE，第 107 卷，2019 年，第 1341 页，第 7 期，第 1302 页。[4] A. Einolghozati、M. Sardari 和 F. Fekri，“基于扩散的分子通信中的中继”，载于 IEEE 信息论国际研讨会（ISIT）论文集，2013 年，第 1844–1848 页。[5] A. Ahmadzadeh、A. Noel 和 R. Schober，“多跳基于扩散的分子通信网络的分析与设计”，IEEE Trans. Mol. Biol. Multi-Scale Commun.，第 1 卷，2015 年，第 157 页，第 2 期，第 144 页。[6] S. Balasubramaniam 等，“基于多跳接合的细菌纳米网络”，IEEE Trans. NanoBiosci.，第 12 卷，2013 年，第 59 页，第 1 期，第 47 页。[7] S. Balasubramaniam 等，“利用细菌特性构建多跳纳米网络”，IEEE Commun. Mag.，第 52 卷，2014 年，第 191 页，第 7 期，第 184 页。[8] R. Zheng 等，“ANIS：用于基于扩散的分子通信的抗噪声符号间干扰抑制滤波器”，IEEE Trans. Mol. Biol. Multi-Scale Co mmun.，第 11 卷，第 4 期，第 572–587 页，2025 年 6 月。[9] Y. Sun 等，“更新还是等待：如何保持数据新鲜”，IEEE Trans. Inf. Theory，第 63 卷，第 11 期，第 7492–7508 页，2017 年。[10] J. T. G´omez 等，“分子通信信道中的信息年龄”，数字信号处理，第 124 卷，第 103108 页，2022 年 5 月。[11] J. Lou 等，“基于多信道的多跳无线网络中的信息年龄优化”，IEEE Trans. Mobile Comput.，第 22 卷，第 10 期，第 5719–5732 页，2022 年。[12] R. Gao 等，“迈向转发年龄：通过时空图神经网络实现的大规模卫星网络深度强化学习路由机制”，IEEE/ACM Trans. Netw.，2025 年。[13] Z. Cheng 等，“多跳移动扩散分子通信的能源效率分析”，Nano Commun. Netw.，第 26 卷，2020 年，文章编号 100313。[14] Z. Cheng 等，“存在漂移时多跳移动扩散分子通信中中继节点位置与判决阈值的联合优化”，IEEE Trans. Nanobiosci.，第 21 卷，第 3 期，第 454–465 页，2022 年。[15] S. Ansari 和 K. A. Alnajjar，“基于遗传算法优化的多跳基于扩散的分子通信路由技术”，IEEE Access，第 11 卷，第 22 689–22 704 页，2023 年。[16] A. Aijaz，“基于扩散的分子纳米网络中的机会路由”，IEEE Wirel. Commun. Lett.，第 4 卷，第 3 期，第 321–324 页，2015 年 6 月。[17] I. Kadota 和 E. Modiano，“具有随机到达的无线网络中的信息年龄最小化”，IEEE Trans. Mobile Comput.，第 20 卷，第 3 期，第 1173–1185 页，2021 年。[18] R. Zheng 等，“基于DNA的分子通信：用于信道建模与检测的马尔可夫方法”，载于 IEEE 全球通信大会（GLOBECOM）论文集。IEEE，2025 年，第 1196–1201 页。[19] R. Zheng 等，“微流控分子通信的系统建模：一种马尔可夫方法”，载于 IEEE 国际通信会议（ICC）论文集，2026 年，第 1–6 页。[20] R. Zheng 等，“基于DNA的分子通信的马尔可夫状态空间建模与信道表征”，IEEE Internet Things J.，2026 年，提前访问。[21] J. A. Boyan 和 M. L. Littman，“动态变化网络中的分组路由：一种强化学习方法”，载于《神经信息处理系统进展》，第 6 卷，1993 年。