Efficient Binomial Channel Capacity Computation with an Application to Molecular Communication

Efficient Binomial Channel Capacity Computation with an Application to Molecular Communication

Richard D. Wesel*, Emily E. Wesel†, Lieven Vandenberghe*, Christos Komninakis‡ and Muriel Medard§
*Department of Electrical and Computer Engineering, UCLA, Los Angeles, CA 90095-1594, Emails: {wesel,vandenbe}@ucla.edu
†Harvard-Westlake, 3700 Coldwater Canyon, Studio City, CA 91604, Email: ewesel1@hwemail.com
‡Qualcomm, 6455 Lusk Blvd, San Diego, CA, 92121. Email: christos@qti.qualcomm.com
‡Department of Electrical Engineering and Computer Science, MIT, Cambridge, MA, 02139. Email: medard@mit.edu

***Abstract*—This paper develops an efficient method to compute the binomial channel capacity and applies it to the molecular channel. The binomial channel (with parameter n) takes the success probability for a Bernoulli trial as input and produces the number of successes in n trials as output. The input alphabet is the unit interval and the output alphabet is the set of integers from zero to n. Despite the fact that the input alphabet is uncountably infinite the capacity-achieving input distributions turn out to have a small finite support that evolves gracefully as n increases. The ellipsoid algorithm was previously used to compute the binomial channel capacity, but convergence is rather slow even with a well-chosen initial condition. The Dynamic Assignment Blahut-Arimoto (DAB) algorithm starts with the capacity-achieving mass point locations for the $n-1$ case and exploits Csiszàr's Min-Max Capacity Theorem to check convergence and adjust mass point locations to achieve a much faster convergence rate, unlocking the potential for the capacity and corresponding input distribution to be computed for larger values of n.**

I. Introduction

The binomial channel (of parameter n) has a channel law defined by the binomial probability distribution of order n. For each channel use, the input X is the probability of success of a Bernoulli trial. The channel output Y is the number of successes observed during n Bernoulli trials. Thus the channel transition probability law is described as

$$P_{Y|X}^{(n)}(y|x) = \binom{n}{y} x^y (1-x)^{n-y}, \qquad (1)$$

where the possible values of y are the integers zero through n. Alternatively, the channel output could be the ordered list of Bernoulli trial outcomes, but since Y is a sufficient statistic [1] of the outcomes for estimating X, the capacity is the same.

Although the input alphabet of the binomial channel is uncountable, the capacity-achieving input distribution has finite support, requiring at most $n+1$ mass points for the binomial channel of parameter n. This was proven in [2] using Dubin's theorem [3]. See also [4] (Corollary 3 in Chapter 4.5).

Fig. 1 shows these finite-support capacity-achieving distributions for $1 \leq n \leq 50$, with the area of the circle indicating the probability of the mass point. For $n = 1$, The mass points are at zero and one resulting in a noiseless binary channel. Then a mass point is introduced at 0.5 which grows in probability as n increases until it splits. As n increases, these two mass points move away from 0.5, and when they are far enough away, a new mass point is born at 0.5.

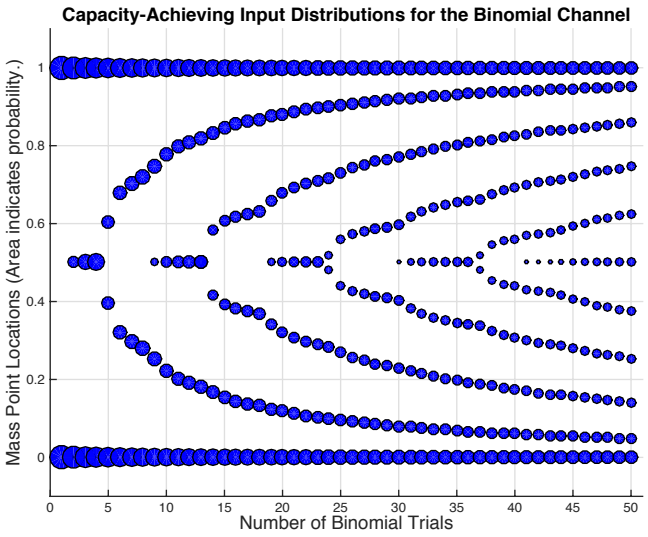

Fig. 1. Capacity-achieving input distributions for the binomial channel with $1 \leq n \leq 50$ obtained by the Dynamic Assignment Blahut-Arimoto algorithm described in Section IV.

The binomial channel has been considered recently as a model for molecular communication [5], [6], where typical values of n will be on the order of a thousand. In [7], values of the capacity of the binomial channel were obtained using the ellipsoid method [8]. However, the ellipsoid method is too slow for practical application when the value of n is large. This paper presents a more efficient approach to computing the capacity of the binomial channel and applies it to the molecular communication channel model of [5].

The rest of the paper proceeds as follows: Section II applies convex optimization to computing the binomial channel capacity. Section III uses the ellipsoid method to solve that problem. Section IV introduces a new method that applies Blahut-Arimoto to a finite support that is dynamically adjusted using Csiszár's Min-Max Capacity theorem. Section V applies this method to the molecular channel model of [5], and Section VI concludes the paper.

II. A Convex Optimization Problem

Despite the fact that finite-support distributions achieve capacity, direct application of the Blahut-Arimoto algorithm [9] is complicated because the locations of the support points within the unit interval are not known. Reasonable approximations can be obtained by applying Blahut-Arimoto with mass points closely spaced along the entire unit interval, with most of these having zero probability. However, we are interested in algorithms that identify the capacity more precisely and also explicitly identify the mass points of the capacity-achieving distribution.

In [7], the capacity of the binomial channel is computed by first formulating the problem as a convex optimization problem and then solving it by using the Ellipsoid method. Assuming that X has discrete support, capacity C_n is

$$\begin{aligned} C_n &= \max_{f_X(x)} I(X;Y) && (2)\\ &= H(Y) - H(Y|X) && (3)\\ &= \max_{f_X(x)} \left\{ H(Y) - \int_{x=0}^{1} f_X(x) H(Y|X=x) f_X(x) dx \right\}. && (4) \end{aligned}$$

Despite the fact that the capacity-achieving distribution on X has at most $n+1$ mass points, the distribution on X is expressed as a density function $f_X(x)$ (and an integral is used in (4)) because the positions of the support points are located anywhere in the uncountable set of the unit interval. Thus, $f_X(x)$ consists of a countable number of delta functions located anywhere in the unit interval.

The optimization problem of (4) can be formulated as a convex optimization problem in a vector space with uncountably infinite dimension. For more mathematical precision, replace $f_X(x)dx$ with $dF(x)$ where $F(x)$ is the cumulative distribution. We allow $dF(x) \in \mathcal{F}$, the set of signed measures on the unit interval and include additional constraints to force $dF(x)$ to be a valid probability distribution. Introducing the additional variables q_y for $y \in \{0,1,\ldots,n\}$ and appropriate equality constraints that force the q_y values to be the output probability distribution $P_Y(y)$ induced by the input distribution yields the following convex optimization **primal problem**:

$$\begin{aligned} \text{minimize} \quad & \sum_{y=0}^{n} q_y \log q_y + \int_{x=0}^{1} f_X(x) H(Y|X=x) dF(x) \\ \text{subject to} \quad & -dF(x) \le 0, \forall x \in [0,1] \\ & \int_{x=0}^{1} dF(x) - 1 = 0\,, \\ & q_y - \int_{x=0}^{1} P_{Y|X}^{(n)}(y|x) dF(x) = 0,\ y \in \{0,\ldots,n\}\,. \end{aligned}$$

The infinite dimensional $dF(x)$ makes the problem intractable. We create a Lagrangian dual problem that can be solved with traditional methods. We introduce Lagrange multipliers $\mu, z_0, z_1, \ldots, z_n$ for the equality constraints and the measurable mapping $v(x)$ of $[0,1]$ to the one-dimensional real space $\mathbf{R}$ for the inequality constraint producing Lagrangian $L(\mathbf{q}, dF(x), v(x), \mathbf{z}, \mu)$ [10]:

$$\begin{aligned} L = & \sum_{y=0}^{n} q_y \log q_y + \int_{x=0}^{1} H(Y|X=x) dF(x) \\ & - \int_{x=0}^{1} v(x) dF(x) + \mu \left(\int_{x=0}^{1} dF(x) - 1 \right) \\ & + \sum_{y=0}^{n} z_y \left(q_y - \int_{x=0}^{1} P_{Y|X}^{(n)}(y|x) dF(x) \right), \end{aligned}$$

which is the cost function augmented with the weighted sum of the constraints. Minimizing $L(\mathbf{q}, dF(x), v(x), \mathbf{z}, \mu)$ with respect to primal variables $\mathbf{q}$ and $dF(x)$ gives the dual function $g(v(x), \mathbf{z}, \mu)$ as follows:

$$g = \inf_{\mathbf{q}, dF(x)} \left\{ \sum_{y=0}^{n} q_y (z_y + \log q_y) - \mu + \int_{x=0}^{1} \gamma(x) dF(x) \right\},$$

where

$$\gamma(x) = H(Y|X=x) - v(x) + \mu - \sum_{y=0}^{n} z_y P_{Y|X}^{(n)}(y|x)\,. \quad (5)$$

Because $dF(x)$ is an unconstrained unsigned measure, $g(v(x), \mathbf{z}, \mu) = -\infty$ unless $\gamma(x) = 0$ for all $x \in [0,1]$, in which case we have

$$g(v(x), \mathbf{z}, \mu) = \inf_{\mathbf{q}} \left\{ \sum_{y=0}^{n} q_y (z_y + \log q_y) - \mu \right\} \quad (6)$$

Setting d/dq_y of the summation in (6) to zero yields the minimizing value of $q_y = \frac{2^{-z_y}}{e}$ so that

$$g(v(x), \mathbf{z}, \mu) = \frac{-\log e}{e} \sum_{y=0}^{n} 2^{-z_y} - \mu\,. \quad (7)$$

The **dual problem** for our primal problem maximizes this $g(v(x), \mathbf{z}, \mu)$ subject to constraints on the slack variables:

$$\begin{aligned} \text{maximize} \quad & \frac{-\log e}{e} \sum_{y=0}^{n} 2^{-z_y} - \mu \\ \text{subject to} \quad & v(x) \ge 0 \quad \forall x \in [0,1], \\ & H(Y|X=x) - v(x) + \mu - \sum_{y=0}^{n} z_y P_{Y|X}^{(n)}(y|x) = 0\,. \end{aligned}$$

Combining these two constraints eliminates the cumbersome infinite-dimensional $v(x)$ producing

$$\begin{aligned} \text{minimize} \quad & \sum_{y=0}^{n} 2^{-z_y} + \frac{\mu e}{\log e} \\ \text{subject to} \quad & H(Y|X=x) + \mu - \sum_{y=0}^{n} z_y P_{Y|X}^{(n)}(y|x) \ge 0 \quad \forall x \in [0,1]\,. \end{aligned}$$

Minimizing the objective function requires the minimum possible value of μ that satisfies

$$\mu \ge \sum_{y=0}^{n} z_y P_{Y|X}^{(n)}(y|x) - H(Y|X=x) \quad \forall x \in [0,1]\,, \quad (8)$$

which leads to the final formulation of the dual problem, in which only the variables z_y remain:

$$\min_{\mathbf{z}} \sum_{y=0}^{n} \frac{2^{-z_y}}{e} + \frac{1}{\log e} \max_{x\in[0,1]} \left\{ \sum_{y=0}^{n} z_y P_{Y|X}^{(n)}(y|x) - H(Y|x) \right\}$$

The dual problem is a finite variable convex optimization problem over the vector $\mathbf{z}$, which can be solved using a variety of techniques.

Once the minimizing $\mathbf{z}$ vector is obtained, complementary slackness indicates that the capacity-achieving mass points are the x values that maximize $\sum_{y=0}^{n} z_y P_{Y|X}^{(n)}(y|x) - H(Y|x)$. The output distribution is recovered using $P_Y(y) = q_y = \frac{2^{-z_y}}{e}$, and the probability $P_X^*(x)$ associated with each mass point can be found by solving the equations

$$P_Y(y) = \sum_{x\in A} P_{Y|X}^{(n)}(y|x) P_X(x) \quad \forall y \in \{0,\ldots,n\}, \tag{9}$$

where A is the set of maximizing x values. The mutual information induced by this P_X is also the capacity.

III. The Ellipsoid Method

In [7], the ellipsoid method was used to solve the the dual problem identified in Section II. The ellipsoid method was developed by Shor, Nemirovski, and Yudin in the 1970's and used by Khachiyan [11] in 1979 to show the polynomial solvability of linear programs. See [8] for an excellent survey. One conclusion of [8] is that the ellipsoid method, while of academic interest, is often not the fastest way to solve a convex problem and can have stability issues as well. However, it is straightforward to program.

The method begins with an initial ellipsoid $\mathcal{E}^{(0)} \in \mathbb{R}^N$ centered at a point z_0 , which is defined as

$$\mathcal{E}^{(0)} = \left\{ z \in \mathbb{R}^N : (z-z_0)^T P_0^T (z-z_0) \le 1 \right\}, \tag{10}$$

and is known to contain the optimizing point z^*. At the k^{th} iteration, the point z_k is at the center of the ellipsoid

$$\mathcal{E}^{(k)} = \left\{ z \in \mathbb{R}^N : (z-z_k)^T P_k^T (z-z_k) \le 1 \right\}. \tag{11}$$

To compute the $\mathcal{E}^{(k+1)}$ we need the subgradient $g_{k+1} \in \mathbb{R}^N$ which is a vector that satisfies $g_{k+1}^T(z^* - z_k) \le 0$, so that

$$z^* \in \mathcal{E}^{(k)} \cap \left\{ z : g_{k+1}^T (z - z_k) \le 0 \right\}. \tag{12}$$

The following computations create a new ellipsoid that contains the half-ellipsoid described above:

$$\tilde{g}_{k+1} = \left(\sqrt{g_{k+1}^T P_k g_{k+1}} \right)^{-1} g_{k+1} \tag{13}$$

$$z_{k+1} = z_k - \frac{1}{N+1} P_k \tilde{g}_{k+1} \tag{14}$$

$$P_{(k+1)} = \frac{N^2}{N^2-1} \left(P_k - \frac{2}{N+1} P_k \tilde{g}_{k+1} \tilde{g}_{k+1}^T P_k \right). \tag{15}$$

The ellipsoid method stopping criterion computes $\sqrt{g_k^T P_k g_k}$ which is an upper bound on the error in the objective function. To apply the ellipsoid algorithm to the dual problem of Section II, $N = n+1$ and the subgradient is the vector with elements

$$-\frac{2^{-z_y} - e P_{Y|X}(y|x^*)}{\log e}, \tag{16}$$

where x^* is any value of x that maximizes

$$\sum_{y=0}^{n} z_y P_{Y|X}^{(n)}(y|x) - H(Y|x) \tag{17}$$

for the current set of z_y values. Also needed are z_0 and P_0 that create an initial ellipse that contains z^*. The simplest approach is to select the origin for z_0 and choose P_0 to be the identity scaled by a value that is larger than the square of the optimizing z^*. Figure 2 shows these squared distances for the first 25 values of n, which grow to over 350 by $n = 25$.

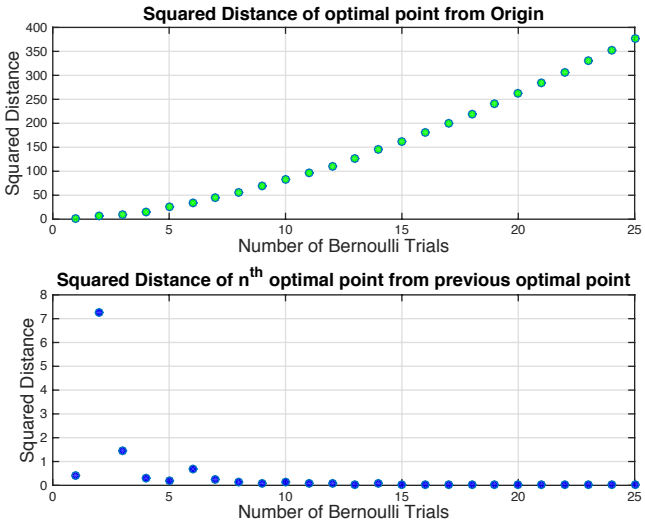

Fig. 2. Squared distance from the origin of optimal points.

However, there is a difficulty in knowing what the squared distance is before the problem has been solved. This problem is avoided by selecting the initial z vector for the $(n+1)^{st}$ case by using the implied by the solution obtained for the n^{th} case as follows:

$$q_y^{(n+1,\text{initial})} = \sum_{x\in A} P_{Y|X}^{(n)}(y|x) P_X^{(n,*)}(x) \quad \forall y \in 0,\ldots,n,$$

$$z_y^{(n+1,\text{initial})} = -\log e q_y^{(n+1,\text{initial})}.$$

In this case, as shown in Figure 2, P_0 can often be the unscaled identity (or the identity scaled by a value less than one). With such a close starting value, one would expect that the ellipsoid method would converge much more quickly. However, as shown in Figure 3, initializing the z vector to the previously optimal point does not significantly improve performance, highlighting the slow convergence of the ellipsoid algorithm even when initialized to a favorable point.

IV. Dynamic Assignment Blahut-Arimoto

Consider again Figure 1. The mass points move only slightly as n progresses, and the only significant changes occur when

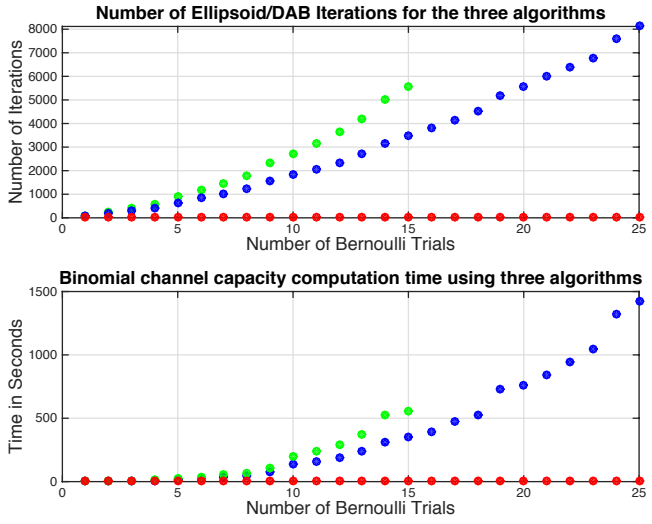

Fig. 3. Number of iterations and computation time in seconds (on a 2.5 GHz MacBook Pro purchased in 2014) running the Ellipsoid and Dynamic Assignment Blahut-Arimoto (DAB) algorithms implemented in Matlab to compute the binomial channel capacity: Ellipsoid method with initial point at the origin (green dots), Ellipsoid method with initial point at position indicated by the solution for the previous n (blue dots), and the DAB algorithm described in Section IV (red dots).

a new mass point is born at $x = \frac{1}{2}$. This section introduces the Dynamic Assignment Blahut-Arimoto (DAB) algorithm as a way to track the gentle evolution of the capacity-achieving distribution as a function of n. At the heart of DAB is Csiszár's Min-Max Capacity Theorem [12], which states:

$$C = \min_{P_Y} \max_{x} D\left(P_{Y|X=x} \| P_Y\right). \tag{18}$$

In fact, we can restate the dual problem found in Section II in terms of Csiszár's Min-Max Capacity Theorem as follows:

$$\min_{P_Y} \left\{ \sum_{y=0}^{n} P_Y(y) - 1 + \max_{x} D\left(P_{Y|X=x} \| P_Y\right) \right\}. \tag{19}$$

The DAB algorithm explicitly identifies the capacity-achieving distribution and capacity of the binomial channel for a sequence of n values. Initialize DAB for a particular n with the mass point locations that achieved capacity for the binomial channel with $n-1$. DAB then performs a series of two-step iterations. At the k^{th} iteration, the first step uses Blahut-Arimoto to optimize the allocation of probability to the current mass point locations to identify $I^{(k)}(X;Y)$ and the maximizing input distribution $P_X^{(k)}$, which is used to compute $P_Y^{(k)}$. The second step computes

$$D_{\max}^{(k)} = \max_{x \in (0,1)} D\left(P_{Y|X=x} \| P_Y^{(k)}\right) \tag{20}$$

$$x_{\max}^{(k)} = \arg\max_{x \in (0,1)} D\left(P_{Y|X=x} \| P_Y\right), \tag{21}$$

and uses these values to check a termination condition and to update the mass point locations if necessary.

Note from (18) that for any P_Y, $D\left(P_{Y|X=x} \| P_Y\right)$ is an upper bound on capacity. $D_{\max}^{(k)}$ is an upper bound on capacity. Regardless of the mass point locations, $I^{(k)}(X;Y)$ is a lower bound on capacity. Thus, DAB computes the difference

$$D_{\max}^{(k)} - I^{(k)}(X;Y) \tag{22}$$

and terminates whenever that difference is within the desired tolerance.

If the termination condition is not met, the mass point locations need to be adjusted so that $D\left(P_{Y|X=x_{\max}^{(k)}} \| P_Y\right)$ is reduced. There are three possible adjustments as follows:

1) If the current number of mass points is even and $x_{\max}^{(k)}$ is closer to 0.5 than any of the mass points, a new mass point is introduced at 0.5.
2) If the current number of mass points is odd and $x_{\max}^{(k)}$ is closest to the mass point at 0.5, then this mass point splits into two mass points $x = 0.5 \pm \delta(x_{\max}^{(k)} - 0.5)$.
3) If neither of the above two conditions is met, then DAB identifies the mass point location x_{closest} that is closest to $x_{\max}^{(k)}$, not including the mass points at zero and one, which never move. This mass point is moved to

$$x_{\text{new}} = x_{\text{closest}} + \delta(x_{\max}^{(k)} - x_{\text{closest}}). \tag{23}$$

Also, the point at location $1 - x_{\text{closest}}$ is moved to $1 - x_{\text{new}}$ preserving symmetry.

Small changes in mass point locations can induce large changes in $D\left(P_{Y|X=x} \| P_Y\right)$. Initially, $\delta = 0.1$ or 0.05 works well, but smaller values of δ are needed as n increases. Note that only one additional mass point is added for any given n.

V. Application to Molecular Channel

In [5] a channel is considered in which imperfect particle-intensity modulation and detection is used for communication between a biological transmitter and receiver (e.g. cells). According to this model, the transmitter may attempt transmission of any number of particles up to a maximum value $m_\tau = \lfloor \lambda \tau \rfloor$, where τ is the symbol duration and λ is the rate at which the transmitter can generate the particles used for communication. The transmitter selects the probability that each of the possible m_τ communication particles is selected for release by the transmitter. In [5] the possible probabilities are limited to fractions $\frac{i}{m_\tau}$ where i is an integer between zero and m_τ. In this paper, we refer to the selection probability as σ and allow it to take on any value in the unit interval.

Not all of the particles selected for release end up being detected at the receiver. Following [5], particles selected for release are *actually* released with probability α. Particles released by the transmitter particle arrive at the receiver with probability ρ, and particles that arrive at the receiver are detected with probability β. The resulting channel is a binomial channel where, by selecting x from the unit interval, the transmitter can effectively select the probability of successful detection to be any value in the closed interval $[0, p_{\max}]$ where $p_{\max} = \alpha\rho\beta$. Considering the input X for this channel to be the aggregate probability of successful detection resulting from the transmitter's choice of selection probability, the channel

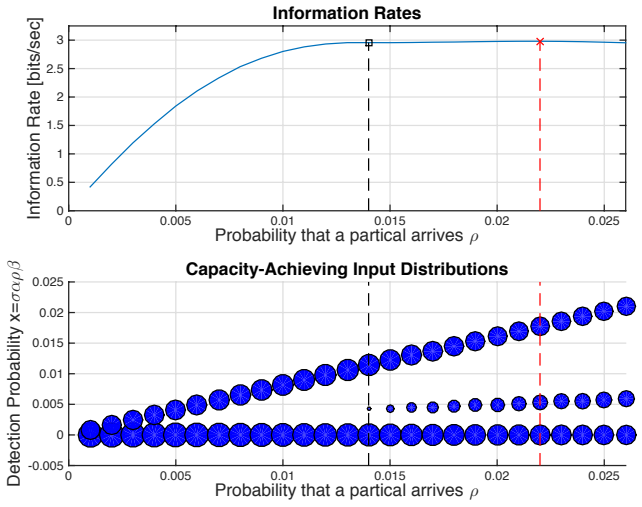

Fig. 4. Information rates and capacity-achieving input distributions as a function of ρ for the binomial channel model of molecular communication from [5] with $c = 1, \eta = 0.2, \alpha = 0.9, \beta = 0.9$ and $\lambda = 1000$. Finite-support capacity-achieving distributions were obtained by the algorithm described in Section IV adapted to allow the maximum value of x to be constrained below 1.

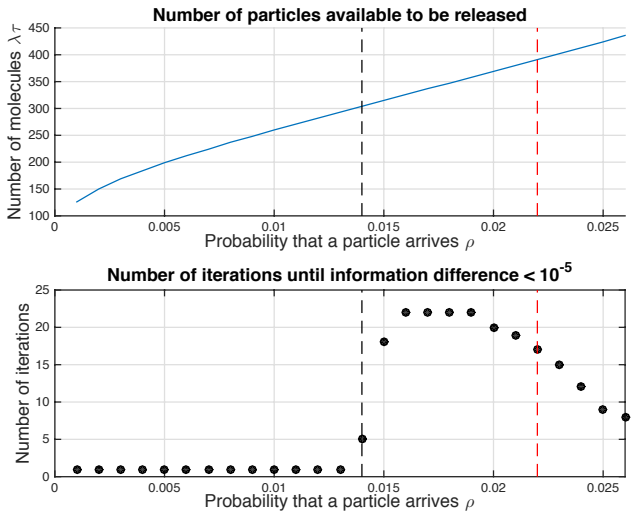

Fig. 5. Number of available particles vs. ρ and number of DAB iterations until the information difference $D\left(P_{Y|X=x}\|P_Y\right) - I(X;Y)$ is below 10^{-5} as a function of ρ for the binomial channel model of molecular communication from [5] with the same parameters and algorithm as in Figure 4.

transition probability law is given by (1) with $n = m_\tau$ and X limited to be in the closed interval $[0, p_{\max}]$.

We will use $C(\tau)$ to refer to the capacity (in bits per channel use), which depends on the symbol time τ through ρ. Larger values of τ increase ρ by allowing more time for particles to arrive at the receiver. Following the model in [5]

$$\tau = \frac{c}{2\text{erfcinv}^2(\rho/\eta)}, \tag{24}$$

with $c = \frac{\ell^2}{2d}$ and $\eta = \frac{r}{\ell+r}$. Here r is the radius of the spherical receiver , d is the diffusion coefficient of particles, and ℓ is the distance between the transmitter and the surface of the spherical receiver.

A key perspective of [5] is to consider the information rate (in bits per second) $\frac{C(\tau)}{\tau}$ rather than bits per channel use and explore the trade-off between the symbol time $\frac{C(\tau)}{\tau}$ and capacity. We use a modified DAB algorithm to compute $\frac{C(\tau)}{\tau}$ for the example case in [5] with $c = 1, \eta = 0.2, \alpha = 0.9, \beta = 0.9$ and λ is 1000. For the scenario of [5] we can use DAB to compute the capacity and optimal mass point locatins for a sequence of ρ values rather than a sequence of n values. The mass points in this modified DAB must lie in the interval $[0, p_{\max}]$, where $p_{\max}$ increases with ρ. There is no longer any assumption of symmetry so that new mass points are introduced at a location between the two middle mass points that depends on $x_{\max}^{(k)}$. The first new mass point is simply introduced at $x_{\max}^{(k)}$. The results are shown in Figures 4 and 5.

VI. Conclusion

The capacity achieving mass points of the binomial channel are relatively small in number and evolve gracefully as n increases. This paper used Csiszár's Min-Max Capacity Theorem to develop the Dynamic Assignment Blahut Arimoto algorithm, a new approach to computing the capacity of the binomial channel and its associated capacity-achieving distribution that is considerably faster than the ellipsoid method that was applied in [7].

References

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, 1991.

[2] H. S. Witsenhausen, "Some aspects of convexity useful in information theory," *IEEE Transactions on Information Theory*, vol. 26, no. 3, pp. 265–271, May 1980.

[3] L. E. Dubins, "On extreme points of convex sets," *Journal of Mathematical Analysis and Applications*, vol. 5, no. 2, pp. 237–244, 1962.

[4] R. G. Gallager, *Informatin Theory and Reliable Communication*. New York: Wiley, 1968.

[5] N. Farsad, C. Rose, M. Medard, and A. Goldsmith, "Capacity of molecular channels with imperfect particle-intensity modulation and detection," May 22 2017, arXiv:1705.08040v1 [cs.IT].

[6] S. Ghavami, R. S. Adve, and F. Lahouti, "Information rates of ask-based molecular communication in fluid media," *IEEE Transactions on Molecular, Biological, and Multiscale Communications*, vol. 1, no. 3, pp. 277–291, September 2015.

[7] C. Komninakis, L. Vandenberghe, and R. D. Wesel, "Capacity of the binomial channel, or minimax redundancy for mempryless sources," in *Proceedings of the IEEE International Symposium on Information Theory*, June 2001.

[8] R. G. Bland, D. Goldfarb, and M. J. Todd, "Feature article—the ellipsoid method: A survey," *Operations Research*, vol. 29, no. 6, pp. 1039–1091, https://doi.org/10.1287/opre.29.6.1039 1981.

[9] R. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Transactions on Information Theory*, vol. 18, no. 4, pp. 460–473, April 1972.

[10] D. G. Luenberger, *Optimization by Vector Space Methods*. John Wiley & Sons, Inc., 1969.

[11] L. G. Khachiyan, "A polynomial algorithm in linear programming," *Doklady Akademiia Nauk SSSR (translated in Soviet Mathematics Doklady*, vol. 20, pp. 191–194, 1979.

[12] I. Csisár and J. Körner, *InformationTheory: Coding Theorems for Discrete Memoryless Systems*, ser. Probability and Mathematical Statistics, Z. Birnbaum and E. Lukacs, Eds. New York - San Francisco - London: Academic Press, 1981. See Theorem 3.4.