

Computing the Length of the Shortest Telomere in the Nucleus

K. Dao Duc and D. Holcman

*Ecole Normale Supérieure, Group of Applied Mathematics and Computational Biology,
 CNRS UMR 8197—INSERM U1024, IBENS, 46 rue d'Ulm 75005 Paris, France*

(Received 25 April 2013; published 27 November 2013)

The telomere length can either be shortened or elongated by an enzyme called telomerase after each cell division. Interestingly, the shortest telomere is involved in controlling the ability of a cell to divide. Yet, its dynamics remains elusive. We present here a stochastic approach where we model this dynamics using a Markov jump process. We solve the forward Fokker-Planck equation to obtain the steady state distribution and the statistical moments of telomere lengths. We focus specifically on the shortest one and we estimate its length difference with the second shortest telomere. After extracting key parameters such as elongation and shortening dynamics from experimental data, we compute the length of telomeres in yeast and obtain as a possible prediction the minimum concentration of telomerase required to ensure a proper cell division.

DOI: [10.1103/PhysRevLett.111.228104](https://doi.org/10.1103/PhysRevLett.111.228104)

PACS numbers: 87.10.Mn, 05.40.-a, 87.16.A-

The ends of the chromosomes, also called the telomeres, are necessary for the maintenance of chromosomal integrity and overall genomic stability [1]. Without any mechanism of elongation, the telomere length can only decrease during cell division [2]. Indeed, DNA polymerase cannot fully replicate telomere ends. As a result, a cell can only divide a finite number of times before it arrests proliferating. This process is called senescence, and has been proposed as a component of cellular aging [3]. Interestingly, ribonucleoprotein telomerase [4] allows telomeres to elongate. Thus, the opposing effects of constitutive shortening and telomerase elongation can maintain a telomere length at equilibrium, so that the cell does not undergo senescence. Short telomeres are preferentially increased by telomerase [5] and the shortest is a limiting factor of cellular proliferation [6]. However, the relation between telomerase activity and the length of the shortest telomeres remains elusive.

Using numerical simulations, previous theoretical models focused on telomere shortening in mammalian cells within a telomerase-deficient cell population and provided significant insights into the molecular dynamics and the variability of triggering senescence [7–10]. More recently, a computational approach based on a deterministic model and stochastic simulations accounted for some experimental distributions of telomeres in senescent cells [11], but did not provide either the underlying physical mechanisms or a study for the distribution of telomeres and the shortest one in particular.

In this Letter, we present a model that describes the dynamics of the telomere length in a telomerase positive cell, with a few specific parameters that we evaluate from experiments in yeast. We model the telomere length by a stochastic jump process, from which we derive the associated jump Fokker-Planck equation. By solving this equation, we show that the telomere distribution is a Gamma

function. In addition, to compute the moments associated with the shortest telomere, we present a statistical analysis for the telomere lengths, which are viewed as independent random variables. We found a universal law for the ratio of the shortest to the second shortest telomeres at steady state. Finally we propose a criterion for telomerase activity to maintain the shortest telomere above a threshold that defines the senescence onset. The present results give a relation between telomerase activity and telomere length and could be further applied to other organisms.

Stochastic model for telomere dynamics during cell division.—We model the length L_n of a single telomere after n divisions. This length is regulated by the opposing effects of constitutive shortening and elongation by the telomerase. Thus, at the end of each cellular division, a telomere can either be elongated by a length ξ with probability $P(L_n)$, which depends on the telomere length, or shortened by a length a with probability $1 - P(L_n)$. After a division, L_{n+1} is then given by

$$L_{n+1} = \begin{cases} L_n - a & \text{with probability } 1 - P(L_n) \\ L_n + \xi & \text{with probability } P(L_n), \end{cases} \quad (1)$$

where a is the shortening length [3–4 base pairs (bp) in yeast [12]], while ξ is the number of nucleotides added by telomerase. Because the number of base pairs added by telomerase does not correlate with telomere length [5], we considered that ξ is a random variable independent of telomere length. Its probability is an exponential distribution [$\Pr(\xi = y) = pe^{-py}$] of parameter p , where $1/p$ is the mean number of base pairs added per elongation. This parameter was obtained by fitting experimental data [5] [see Fig. 1(a)], for which we obtained $p = 0.026$ (mean number of nucleotides added ≈ 40 bp). We further inferred the probability $P(L_n)$ from analyzing biochemical reactions involved in telomerase recruitment (see the Supplemental Material [13]). Indeed, we considered the two opposite

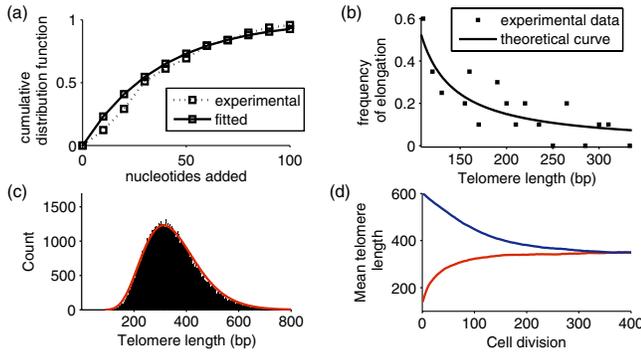


FIG. 1 (color online). Telomere length in the yeast *S. cerevisiae*. (a) CDF of nucleotides added per elongation, fitted to Ref. [5]. (b) Probability of elongation as a function of the telomere length computed from Eq. (2) and fitted to Ref. [5] (dots). (c) Equilibrium telomere length distribution simulated from Eq. (1) ($n = 500, 100\,000$ runs) compared with the analytical stationary distribution (red) computed from Eq. (7). (d) Average dynamics (10 000 runs) of short (150 bp) and long (600 bp) telomere return to steady state.

processes of elongation due to telomerase activity with a rate k_2 , and inhibition of telomerase, the rate of which depends on telomere length [14,15], and which is given by $k_1(L_n - L_0)$. Combining these two opposite reactions, we obtain for the probability of elongation

$$P(L_n) = \begin{cases} \frac{1}{1 + \beta(L_n - L_0)} & \text{if } L_n > L_0 \\ 1 & \text{else,} \end{cases} \quad (2)$$

where L_0 is a characteristic telomere length, and $\beta = k_1/k_2$ is the effective telomerase inhibition rate (when β increases, the probability of elongation decreases). We fitted expression (2) to yeast data [see Fig. 1(b)] and we found $\beta = 0.045 \text{ bp}^{-1}$ and $L_0 = 90 \text{ bp}$. Interestingly, the present model is in agreement with previous observations [14] that after a cell division, the mean change of telomere length $\langle \Delta(L) \rangle$ is a decreasing function, larger for short telomeres than for long ones. Indeed from Eq. (1) of our model, we derive that $\langle \Delta(L) \rangle = -a[1 - P(L)] + (1/p)P(L) = -a + [(a + 1/p)/(1 + \beta(L - L_0))]$. Using the numerical values for p , β , and L_0 in Eq. (1), we simulate a population of telomere lengths and obtain a steady state distribution [see Fig. 1(c)] with mean 350 ± 102 (standard deviation) base pairs, in agreement with experimental data in yeast [16]. Finally, we simulate the return to steady state for average telomere length [see Fig. 1(d)] with different initial conditions corresponding to either short (150 bp) or long (600 bp) telomeres (similar simulations were also reported in mammalian cells [11]). The resulting dynamics is similar to experimental measurements [14] with the following quantification. Starting from a short telomere length, we observe a linear convergence ($\mu = \lim_{n \rightarrow \infty} [|L_n - L_\infty|/|L_{n-1} - L_\infty|] < 1$) to equilibrium with a rate μ of 0.98 (simulation) compared to 0.93 (experiments) [14], while starting from long

telomeres the dynamics is divided into two phases: an initial constant shortening length of 1.9 bp (simulation) versus 2.1 bp (experiments) [14], followed by an exponential decay [see Fig. 1(d)]. We conclude that the present stochastic model captures the average dynamics and allows us to reproduce the experimental distribution of telomeres in yeast, confirming its validity. We shall now use this model to compute the distribution of telomeres, some statistical quantities associated with the shortest telomere and to obtain novel predictions.

The stationary distribution of telomere length.— We compute here the steady state distribution based on Eq. (1). Because the shortening length (3.5 bp in yeast) is small compared with the elongation ($\approx 40 \text{ bp}$ on average), we rescale Eq. (1) with a time scale Δt and normalize the telomere length L using the scaled variable $X = (L - L_0)/a$. Equation (2) becomes the jump rate function $\lambda(X) = 1/(1 + BX)$, and the probability for the jump ξ is given by $\Pr(\xi = y) = \theta e^{-\theta y} = b(y)$, where $B = a\beta$ (elongation probability parameter) and $\theta = ap$ (elongation length parameter) are parameters of the normalized model. In the limit $\Delta t \rightarrow 0$, we obtain a continuous process with a constant drift for shortening, and possible large jumps with exponential rates for elongation, and the dynamics of the normalized length X is given by

$$\begin{aligned} X(t + \Delta t) &= \begin{cases} X(t) - \Delta t & \text{with probability } 1 - \lambda(X)\Delta t + o(\Delta t) \\ X(t) + \xi & \text{with probability } \lambda(X)\Delta t + o(\Delta t). \end{cases} \end{aligned} \quad (3)$$

Since the mean elongation length $1/p$ ($\approx 40 \text{ bp}$) is large compared with the shortening ($\approx 3 \text{ bp}$ per division) and the jump rate is close to 1 when X is close to 0, we neglected the probability that $X(t)$ goes to negative values. Otherwise, a zero boundary condition has to be imposed on the probability density function (pdf) of X at zero. The pdf $f(x, t) = \partial F(x, t)/\partial x$ where $F(x, t) = \Pr\{X(t) \leq x\}$ as $\Delta t \rightarrow 0$ satisfies the Takacs equation, which is the forward Fokker-Planck equation for Eq. (3) and $x > 0$, [17,18]

$$\partial_t f = \partial_x f - \lambda(x)f(x, t) + \int_0^l \lambda(y)f(y, t)b(x - y)dy. \quad (4)$$

The stationary distribution function \bar{f} satisfies

$$0 = \bar{f}'(x) - \lambda(x)\bar{f}(x) + \int_0^x \lambda(y)\bar{f}(y)b(x - y)dy, \quad (5)$$

with normalization condition $\int_0^\infty \bar{f}(x)dx = 1$. The general solution of Eq. (5) is [see the Supplemental Material [13]]

$$\begin{aligned} \bar{f}(x) &= L_1(1 + \theta x)e^{-\theta x} \mathcal{M}\left(\frac{-1}{B} + 1, \frac{-1}{B} + 2, \theta\left(x + \frac{1}{B}\right)\right) \\ &\quad + L_2(Bx + 1)^{1/B} e^{-\theta x}, \end{aligned} \quad (6)$$

where L_1 and L_2 are two constants and \mathcal{M} is the Kummer confluent hypergeometric function. The asymptotic

behavior of $\mathcal{M}(a, b, z)$ for $z \rightarrow +\infty$ is $\mathcal{M}(a, b, z) \approx (e^z a^{a-b})/\Gamma(a)$. Thus $L_1 = 0$ (otherwise, $\lim_{x \rightarrow \infty} \bar{f}(x) = \lim_{x \rightarrow \infty} [(L_1 B x e^{p/B})/(1 - 1/B)\Gamma(1 - 1/B)] = \pm\infty$, and \bar{f} would not be a probability function), and L_2 is determined by the normalization condition. We conclude that the steady state distribution for the normalized telomere length is

$$\bar{f}(x) = \frac{\theta[\theta(x + \frac{1}{B})]^{1/B} e^{-\theta[x+(1/B)]}}{\Gamma(\frac{1}{B} + 1, \frac{\theta}{B})}, \quad (7)$$

where $\Gamma(s, x) = \int_x^{+\infty} t^{s-1} e^{-t} dt$ is the upper incomplete Gamma function. Finally, the variable $\theta[X + (1/B)]$ follows approximatively a normalized Gamma distribution of parameter $\alpha = 1 + (1/B)$ [Eq. (7)]. The mean $\langle L \rangle$ and standard deviation (SD) for the telomere length $L = aX + L_0$ are given by [see the Supplemental Material [13]]

$$\langle L \rangle = L_0 - \frac{1}{\beta} + \frac{1}{p} \left(1 + \frac{1}{a\beta}\right) \quad (8)$$

$$\text{SD}(L) = \frac{\sqrt{1 + \frac{1}{a\beta}}}{p}. \quad (9)$$

In Figs. 1(c) and 2, we obtain a good agreement between the probability density function $\Gamma(\alpha, 1)$ and the empirical distributions obtained from stochastic simulations [Eq. (1)]. The distributions are robust for the parameters fitted in yeast [see Fig. 1(c)] as for other values of θ (elongation length parameter) and B (elongation probability parameter) (see Fig. 2). In particular, changing the parameter θ resulted only in a shift of the length distribution. In contrast, changing the elongation probability parameter B (which accounts for the telomerase activity) affects significantly the shape of the distribution through the parameter α (see Fig. 2), suggesting that the level of telomerase can drastically modify the telomere length distribution.

The distribution of the shortest telomere.—Because the shortest telomeres potentially initiate senescence signaling [6], we shall now focus on the distribution of the shortest telomere in an ensemble of $2n$ telomeres, which

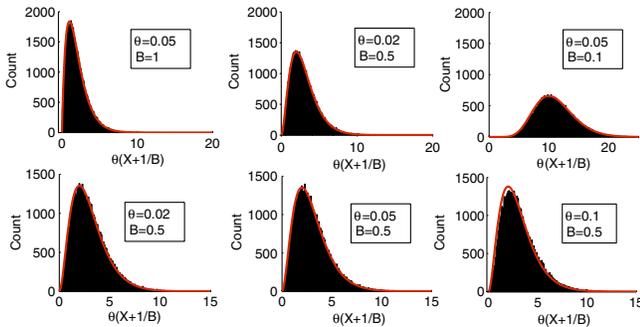


FIG. 2 (color online). Stationary distribution of telomere length for various values of θ and B . Histograms of $\theta(X_n + 1/B)$ using Eq. (1) (500 steps, 50 000 runs), compared with the scaled Gamma pdf of parameter $1 + (1/B)$.

correspond to n chromosomes (16 in yeast and n is in the range of 36–60). We model the $2n$ telomere lengths as independent identically distributed variables L_1, L_2, \dots, L_{2n} . Considering $2n$ independent identically distributed variables X_1, \dots, X_{2n} following a distribution f , the pdf of $X_{(1:2n)} = \min(X_1, X_2, \dots, X_{2n})$ is given by

$$f_{X_{(1:2n)}}(x) = 2n(1 - F)^{2n-1}(x)f(x), \quad (10)$$

where $F(x) = \int_0^x f(u)du$. The statistical moments $\bar{X}_{(1:2n)}^k = \int_{\mathbb{R}^+} x^k f_{X_{(1:2n)}}(x)dx$ are given by

$$\bar{X}_{(1:2n)}^k = k \int_{\mathbb{R}^+} x^{k-1} (1 - F)^{2n}(x)dx. \quad (11)$$

When f is a Gamma distribution of parameter α and n is sufficiently large, we estimate $\bar{X}_{(1:2n)}^k$ using Laplace's method. Equation (7) with $\alpha = 1 + (1/B)$ is such that, when x goes to 0, $F(x) \approx mx^r$ with $m = [1/\alpha\Gamma(\alpha)] > 0$ and $r = \alpha > 1$ (see also p. 305 of Ref. [19]), we obtain

$$\bar{X}_{(1:2n)}^k \approx \frac{k}{r} \int_0^{+\infty} x^{k/r-1} \exp[2n \ln(1 - F(x^{1/r}))] dx.$$

Thus, $\bar{X}_{(1:2n)}^k \approx [(k\Gamma(k/r))/(r(2nm)^{k/r})]$. That is,

$$\bar{X}_{(1:2n)}^k \approx \frac{k\Gamma(\frac{k}{\alpha})}{\alpha} \left(\frac{\alpha\Gamma(\alpha)}{2n}\right)^{k/\alpha}. \quad (12)$$

Using formulas (10)–(12), we compute the pdf and the moments of the shortest telomere length $L_{1:2n}$ for $k = 1$. As shown from Eq. (7), we have that $p[(L_i - L_0)/a + (1/B)]$ follows a Gamma distribution. Then using Eq. (12), we obtain that the mean of the shortest telomere length is

$$\bar{L}_{(1:2n)} \approx L_0 - \frac{1}{B} + \frac{\Gamma(1 + \frac{1}{B})\Gamma(\frac{1}{1+1/B})^{1/(1+1/B)}}{p(1 + \frac{1}{B})^{1/(1+B)}(2n)^{1/(1+1/B)}}. \quad (13)$$

In Fig. 3(a), we obtain for various values of the elongation probability parameter B ($n = 50$) a good agreement between the analytical formula (12) and expression (11) that was computed from the empirical approximation of Eq. (7) ($B > 0.5$). We conclude that the shortest telomere length can be obtained from formula (12) with B large enough (> 0.5). When B is small, Laplace's method fails to approximate well the cumulative distribution F , and underestimates $\bar{X}_{(1:n)}$. Using values $p = 0.026$, $\beta = 0.045$ ($B = 0.16 < 0.5$), and $L_0 = 90$ (yeast) and Eq. (11) for the first two moments ($k = 1$ and 2), we compute that the mean shortest telomere length is 184 ± 25 bp (confirmed by the empirical simulations of Eq. (1) [see Fig. 3(b)]). Thus the shortest telomere length should explain variations of senescence onset observed in telomerase-deficient cells [20,21].

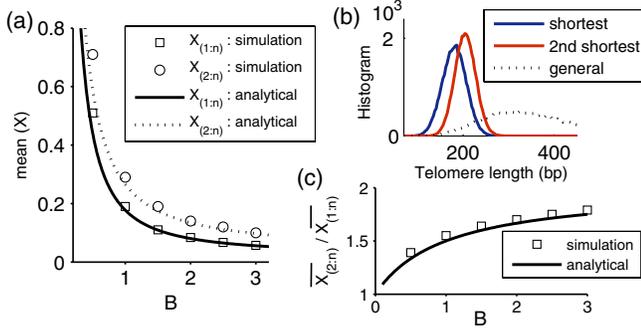


FIG. 3 (color online). Shortest telomeres. (a) Mean shortest $X_{(1:2n)} = \theta(L_1 + 1/B)$ (circle) and second shortest telomere $X_{(2:2n)} = \theta(L_2 + 1/B)$ scaled lengths (square) as a function of B ($\theta = 0.02$, $n = 50$, runs = 10000), compared with formula (12) (continuous line) and formula (16) (dotted line). (b) Simulated distribution of the two shortest telomere lengths L_1 and L_2 among 32 (100000 runs) with the distribution of Fig. 1(c) (dotted line). (c) Comparing the ratio $\bar{X}_{(2:2n)} / \bar{X}_{(1:2n)}$ computed from formula (16) (continuous line) and numerically (square).

The gap between the shortest telomere and the others.— We study here the separation between the shortest telomere and the others. We compute the distribution of the second shortest length $X_{(2:2n)}$. The pdf $f_{X_{(2:2n)}}$ of $X_{(2:2n)}$ is given by

$$f_{X_{(2:2n)}}(x) = 2n(2n-1)F(x)(1-F(x))^{2n-2}f(x), \quad (14)$$

and the statistical moments $\bar{X}_{(2:2n)}^k$ satisfy the induction relation

$$\bar{X}_{(2:2n)}^k = n\bar{X}_{(1:2n-1)}^k - (2n-1)\bar{X}_{(1:2n)}^k. \quad (15)$$

Using Eq. (12) for $k = 1$, we obtain the surprising result that the ratio $\bar{X}_{(2:2n)} / \bar{X}_{(1:2n)}$ for n or $B \gg 1$ asymptotically depends on $\alpha = 1 + (1/B)$ as

$$\frac{\bar{X}_{(2:2n)}}{\bar{X}_{(1:2n)}} \approx 1 + \frac{1}{\alpha}. \quad (16)$$

Interestingly, in the generic case where the pdf of random variables has a nonzero first order derivative at 0, this ratio is universal and equal to $\frac{3}{2}$. In Fig. 3(c), we plot Eq. (16) and obtain good agreement with numerical simulations, which confirms the accuracy of the analytical approach for estimating the gap between the shortest telomere and the others. In the case of the yeast, we use Eq. (15) and we obtain that the mean length of the second shortest telomere is 207 bp [confirmed by Fig. 3(b)]. Thus, we predict that the shortest telomere is on average 22 bp shorter than the second one.

Maintenance of telomere length.—As the size of the shortest telomere is controlled by the elongation probability parameter B (which depends on the telomerase concentration), we propose to estimate now the minimum value B_c for which the shortest telomere does not go below a critical length with probability 99%. The critical length L_q

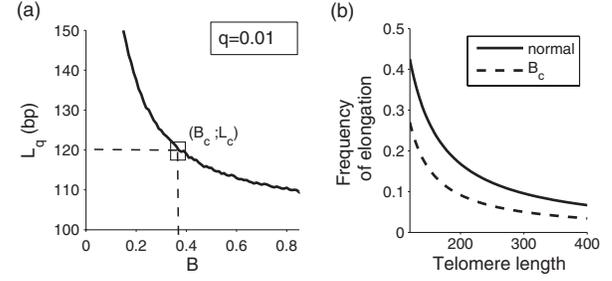


FIG. 4. Probability to maintain the shortest telomere over a critical length. (a): We plot the length L_q for which $P(L_{1:n} < L_q) = q$, as a function of B , with $q = 0.01$. We indicate the critical value B_c for which $L_q = L_c = 120$ bp. (b) Frequency of elongation for normal value of B and B_c obtained using Eq. (17).

is defined by $P(L_{(1:2n)} < L_q) = q$ which is the probability that the shortest telomere length is less than L_q with probability $q = 0.01$. Using that the dimensionless length $X = (L - L_0)/a$ and $\theta(X + (1/B))$ is a Gamma distribution of parameter $1 + (1/B)$ [Eq. (7)], we obtain that $L = L_0 - [1/B + (aY/\theta)]$, where Y is again a Gamma distribution of parameter $1 + (1/B)$. Using Eq. (10), we get that L_q is such that

$$1 - \left[1 - F_Y\left(\left(L_q + \frac{1}{B} - L_0\right)\frac{B}{a}\right) \right]^{2n} = q, \quad (17)$$

where $Y \sim \Gamma(1 + (1/aB), 1)$. A direct computation gives L_q as a function of B that we plotted in Fig. 4(a). For a threshold $L_c = 120$ bp below which the telomerase cannot regulate telomere length (an event that triggers senescence [5]), we find that $B_c = 0.36$ [see Fig. 4(a)]. For this critical value, the frequency of elongation [Eq. (2)] in Fig 4(b) is roughly divided by 2 in the range of 100–300 bps and the probability of senescence for a single cell after one division is 1%. In contrast, with the endogenous value $\beta = (B/a) = 0.045$ we compute that this probability is quasi-null, as it drops to $1 - (1 - 1.5 \times 10^{-40})^{32} \approx 5 \times 10^{-38}$.

In conclusion, we presented a stochastic approach to compute the telomere length distribution. We specifically focused on the shortest one. This approach allowed us to find a condition on the parameter β so that the length of the shortest telomere does not decrease below a critical length after a division, with a probability of 1%. It would be interesting to relate this measurable parameter β to the level of telomerase, as this relation is still unclear [11]. Another interesting extension is the modification of β in cancerous cells, where the cell cycle duration is affected [22,23]. Since such changes might increase the telomerase effectiveness, this could indeed explain their intrinsic proliferation ability.

This research is supported by an ERC Starting Grant. We thank Teresa Teixeira and Zhou Xu for fruitful discussions. Study [24] provides complementary information to the one presented here with genetics experiments.

- [1] T. de Lange, L. Shiue, R. M. Myers, D. R. Cox, S. L. Naylor, A. M. Killery, and H. E. Varmus, *Mol. Cell Biol.* **10**, 518 (1990).
- [2] J. D. Watson, *Nat. New Biol.* **239**, 197 (1972).
- [3] G. Aubert and P. M. Lansdorp, *Physiol. Rev.* **88**, 557 (2008).
- [4] T. R. Cech, *Cell* **116**, 273 (2004).
- [5] M. T. Teixeira, M. Arneric, P. Sperisen, and J. Lingner, *Cell* **117**, 323 (2004).
- [6] M. T. Hemann, M. A. Strong, L. Y. Hao, and C. W. Greider, *Cell* **107**, 67 (2001).
- [7] Z. Tan, *J. Theor. Biol.* **198**, 259 (1999).
- [8] C. J. Proctor and T. B. Kirkwood, *Mech. Ageing Dev.* **123**, 351 (2002).
- [9] C. J. Proctor and T. B. Kirkwood, *Aging Cell* **2**, 151 (2003).
- [10] J. op den Buijs, P. P. van den Bosch, M. W. Musters, and N. A. van Riel, *Mech. Ageing Dev.* **125**, 437 (2004).
- [11] I. A. Rodriguez-Brenes and C. S. Peskin, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 5387 (2010).
- [12] M. S. Singer and D. E. Gottschling, *Science* **266**, 404 (1994).
- [13] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevLett.111.228104> for a model to compute the probability of elongation, the detailed solution of the Takacs equation, and the variance of the telomere length.
- [14] S. Marcand, E. Gilson, and D. Shore, *Science* **275**, 986 (1997).
- [15] A. J. Lustig, S. Kurtz, and D. Shore, *Science* **250**, 549 (1990).
- [16] R. J. Wellinger and V. A. Zakian, *Genetics* **191**, 1073 (2012).
- [17] Z. Schuss, *Diffusion and Stochastic Processes: An Analytical Approach* (Springer, New York, 2010).
- [18] L. Kleinrock, *Queueing Systems* (Wiley, New York, 1975), Vol 1.
- [19] K. Lange, *Applied Probability* (Springer, New York, 2003).
- [20] V. Lundblad and J. W. Szostak, *Cell* **57**, 633 (1989).
- [21] K. B. Ritchie, J. C. Mallory, and T. D. Petes, *Mol. Cell Biol.* **19**, 6065 (1999).
- [22] S. Cos, J. Recio, and E. J. Sánchez-Barceló, *Life sciences* **58**, 811 (1996).
- [23] X. Wu, J. A. Roth, H. Zhao, S. Luo, Y. L. Zheng, S. Chiang, and M. R. Spitz, *Cancer Res.* **65**, 349 (2005).
- [24] Z. Xu, K. D. Duc, D. Holcman, M. T. Teixeira, *Genetics* **194**, 847 (2013).