

De biologische cel als minicomputer

Het multiplexen van biochemische signalen

We beschrijven een eenvoudig cellulair netwerk dat tegelijkertijd meerdere signalen door een gedeelde signaalcascade kan sturen. Zowel de codering als decodering voor dit proces zijn niet triviaal vanwege de mogelijke overspraak tussen de signalen. We laten zien dat het beschreven netwerk 100% betrouwbaar is, zelfs in de aanwezigheid van intrinsieke netwerkkruis.

Wiet de Ronde en Pieter Rein ten Wolde

Een kosten-batenafweging

Biologische cellen leven in een omgeving waarin talrijke verschillende signalen aanwezig zijn, zoals de temperatuur of het suikeraanbod. Om goed te functioneren moeten cellen gepast op deze signalen reageren. Organismen kunnen daarom als informatie-verwerkende machines beschouwd worden. Zelfs simpele organismen kunnen een groot aantal verschillende signalen waarnemen, doorsturen en vervolgens hun gedrag of fenotype aanpassen. Het doorgeven van informatie gebeurt met zogenaamde signaaltransductienetwerken, waarin eiwitten fysisch en chemisch interageren.

Een grote uitdaging voor elke cel is om de informatie in ieder signaal

betrouwbaar te verwerken. Dit komt omdat op cellulair niveau signaalcascades worden gekenmerkt door toevalsprocessen. De componenten van een cascade – de eiwitten – bewegen zich vaak op een willekeurige manier door de cel op zoek naar elkaar. Daarnaast zijn de reacties stochastisch van aard: chemische en fysische reacties zijn kansprocessen.

Een andere belangrijke uitdaging voor de cellen is het enorme aantal verschillende signalen. Immers, als voor elk signaal een aparte signaalcascade zou bestaan, dan zou de cel een groot aantal verschillende eiwitten moeten produceren, wat een energetisch kostbare zaak is. Recent hebben experimenten echter aangetoond dat signaaltransductienetwerken componenten vaak delen [1,2]. Dit bracht ons tot de vraag of cellen biochemische signalen kunnen ‘multiplexen’: kunnen ze meerdere signalen door een en hetzelfde signaaltransductienetwerk sturen, net zoals meerdere telefoonsignalen door dezelfde koperdraad of glasvezelkabel gestuurd kunnen worden?

Zonder multiplexing zou telecommunicatie zoals we die vandaag de dag gebruiken simpelweg niet mogelijk zijn: er zou dan immers een aparte lijn tus-

sen elk paar telefoons op aarde aangelegd moeten worden. Het gevaar bij multiplexen is echter dat de verschillende signalen over de lijn elkaar beïnvloeden, dat wil zeggen crosstalk of overspraak genereren. Om dit te voorkomen is een betrouwbaar multiplexschema noodzakelijk. In dit artikel bespreken we een simpel én betrouwbaar schema voor het multiplexen van biochemische signalen.

Multiplexen is een mappingprobleem

We beschrijven een proof-of-concept-model. We beschouwen daarom het eenvoudigste netwerk dat twee onafhankelijke signalen, S_1 en S_2 , kan multiplexen. Elk signaal heeft een eigen,

30

Wiet de Ronde (1981) studeerde technische natuurkunde aan de Technische Universiteit Delft. Daarna promoveerde hij in de groep Biochemical Networks van Pieter Rein ten Wolde aan het FOM Instituut AMOLF, waarbij hij zich heeft gericht op de betrouwbaarheid en de efficiëntie van signaalcascades in cellen.



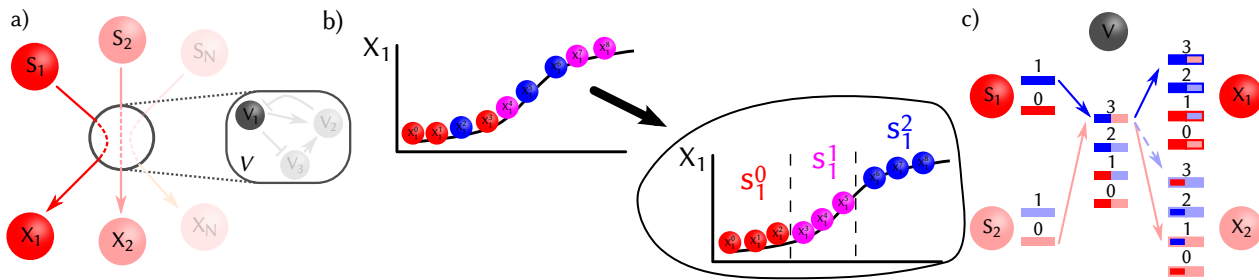
deronde@amolf.nl

Pieter Rein ten Wolde (1970) studeerde fysische chemie aan de Rijksuniversiteit Groningen. Zijn promotieonderzoek verrichtte hij op het FOM Instituut



AMOLF onder begeleiding van Daan Frenkel. Na een postdoc in de groep van David Chandler aan de University of California in Berkeley startte hij in 2001 de onderzoeksgroep Biochemical Networks op AMOLF.

tenwolde@amolf.nl



Figuur 1 a) Schematische weergave van het model. Dun weergegeven is de generieke extensie. b) Om de multiplexvoorwaarde te illustreren, beschouwen we hier tweeingangssignalen, die elk drie toestanden hebben. Dit betekent dat zowel X_1 als X_2 elk negen toestanden heeft, corresponderend met de 3×3 mogelijke combinaties vaningangssignalen. Bij één toestand van hetingangssignaal S_1 horen dus drie toestanden van X_1 , overeenkomend met de drie verschillende toestanden van X_2 ; de drie toestanden van X_1 die horen bij één toestand van S_1 zijn aangegeven met dezelfde kleur. De multiplexvoorwaarde is dat i) de drie toestanden van X_1 die horen bij één toestand van S_1 zijn gegroepeerd en ii) de verschillende groepen elkaar monotoon opvolgen. c) Correcte mapping van de twee toestanden in S_1 en S_2 naar vier toestanden in V en X_1, X_2 ; er wordt voldaan aan de multiplexingvoorwaarde (bijvoorbeeld de toestanden x_1^2, x_1^3 corresponderen met s_1^1 .)

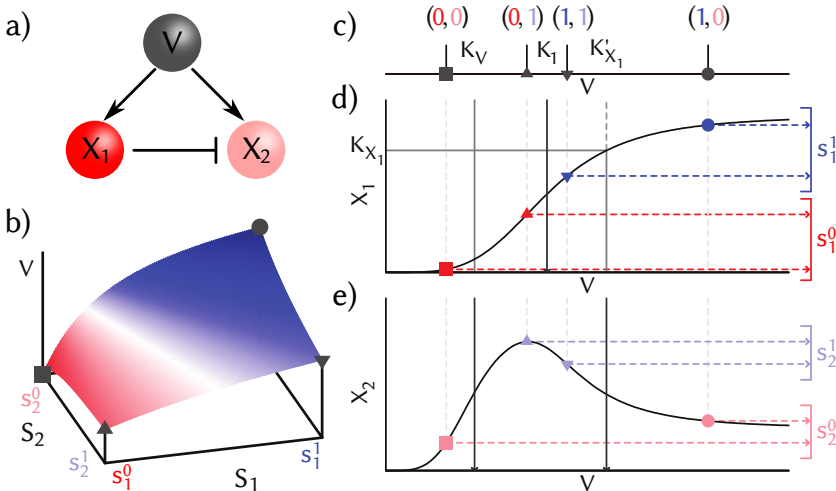
unieke respons (X_1 voor S_1 en X_2 voor S_2) en de signalen delen een signaal-cascade, de component (V) (figuur 1a). We nemen aan dat elk signaal in een van twee toestanden verkeert, aangeduid respectievelijk met τ (bijvoorbeeld s_1^1) en o . Dit kan bijvoorbeeld betekenen dat een signaal wel (τ) of niet (o) aanwezig is. Gecombineerd zijn er dus vier verschillende ingangstoestanden. Deze vier ingangstoestanden worden gemapped op V , zodat voor V vier toestanden bestaan (v^0, v^1, v^2 en v^3), en vervolgens op de responsen X_1 en X_2 , zodat ook deze elk vier toestanden hebben (figuur 1c). De signalen S_1 en S_2 zijn dus binair, terwijl V, X_1 en X_2 meer dan twee toestanden hebben.

Voor de cel is het belangrijk dat ze met de kennis van de toestand van een uitgangssignaal (bijvoorbeeld X_1) kan bepalen wat de toestand van het corresponderendeingangssignaal is. Dit betekent dat de twee toestanden van X_1 die horen bij s_1^1 gegroepeerd moeten zijn, net zoals de twee toestanden die horen bij s_1^0 (figuur 1b). Daarnaast moeten de verschillende groepen elkaar monotoon opvolgen (figuur 1b). Alleen dan is het mogelijk om, zonder complexe bijkomende regeling, aan de hand van de respons x_1^i direct te bepalen wat de toestand van S_1 is. Dit kan bijvoorbeeld als de twee lage toestanden x_1^0, x_1^1 corresponderen met s_1^0 , terwijl x_1^2, x_1^3 de hoge toestand s_1^1 signaleren. De combinatie van deze twee eisen noemen we de ‘multiplexingvoorwaarde’. In figuur 1c tonen we een mapping die hieraan voldoet.

Tot nu toe hebben we niet beschreven wat de ‘toestand’ van een component precies betekent. In het netwerk dat we zullen beschrijven is de toestand

A	s_2^0	s_2^1	B	s_2^0	s_2^1	C	s_2^0	s_2^1
s_1^0	0	2	s_1^0	0	1	s_1^0	0	1
s_1^1	3	1	s_1^1	2	3	s_1^1	3	2

Figuur 2 De drie verschillende mappings A, B en C. De ordening van de toestanden in V is niet triviaal voor mapping A en C.



Figuur 3 a) De incoherente feed-forward lus; V activeert X_1 en X_2 , en X_1 onderdrukt X_2 . b) De input-outputrelatie tussen V en S_1, S_2 . c) De vier verschillende toestanden in V corresponderen met de vier ingangstoestanden voor S_1 en S_2 . d) Activatie van X_1 door V is voldoende voor correcte decoding. e) De negatieve regulatie door X_1 zorgt voor de niet-monotone respons van X_2 als functie van V . Kleuren als in figuur 1c.

gelijk aan het aantal kopieën, of de concentratie, van een eiwit. Echter, dit is slechts een keuze; de toestand kan bijvoorbeeld ook de modificatietoestand of oscillatiefrequentie van een bepaald eiwit zijn [3].

Decoding van V naar X_1 en X_2

Het is instructief om eerst het decoderprobleem van V naar X_1 en X_2 te begrijpen alvorens de codering van S_1 en S_2 in V te beschrijven. De vier concentraties (toestanden) van V (v^0, v^1, v^2 en v^3) corresponderen ieder met een

van de vier ingangstoestanden. Omdat verwisseling van S_1 en S_2 en verwisseling van de o - en τ -toestand leiden tot symmetrisch gelijke mappings, kunnen deze vier ingangstoestanden slechts op drie unieke manieren worden gemapped in het concentratieniveau van V (figuur 2a-c).

Is het mogelijk om met een eenvoudig biochemisch netwerk een mapping te decoderen zodanig dat wordt voldaan aan de bovenstaande multiplexingvoorwaarde? Met alléén directe regulatie van V naar X_1 en X_2 is dit niet

Mathematisch model

Ons netwerk bestaat uit twee componenten: codering en decodering.

Codering

De codering beschrijven we met een Monod-Wyman-Changeuxmodel (MWC-model) [6]. Dit statistisch-fysisch model beschrijft het gedrag van een eiwit (bijvoorbeeld V) dat in twee conformaties kan voorkomen; 'actief' (A) en 'inactief' (I). In dit model wordt van elke mogelijke conformatietoestand van V aan de hand van de vrije energie (E) de waarschijnlijkheid bepaald. De kans op de actieve conformatie p_V^A volgt dan uit de verhouding van alle conformatietoestanden waarin V actief is met alle mogelijke conformatietoestanden van V. Voor twee toestanden is de berekening eenvoudig. We beschouwen de inactieve toestand als de referentietoestand ($E_0 = 0$), $\omega_0 = e^{-E_0/k_B T} = 1$, en de kans op de actieve toestand p_V^A wordt dan bepaald door de vrije energie van de actieve toestand via $\omega_A = e^{-E_A/k_B T}$, met k_B de Boltzmannconstante en T de temperatuur:

$$p_V^A = \frac{\omega_A}{\omega_A + \omega_0}. \quad (1)$$

Als eiwitten (bijvoorbeeld S_1 , S_2) aan V kunnen binden, is het aantal mogelijke toestanden van V groter, omdat V tegelijkertijd actief/inactief en gebonden/ongebonden kan zijn. De kans op een bepaalde toestand hangt af van de conformatie van V en de bindingsenergie. De bindingsenergie is een functie van de concentratie en de dissociatieconstante K_D . K_D verschilt per eiwit en kan afhankelijk zijn van de structuur (de conformatie) van V. Dit

leidt voor p_V^A tot

$$p_V^A = \frac{\omega_A \left(\omega_0 + \frac{S_1}{K_{D,A}^{S_1}} + \frac{S_2}{K_{D,A}^{S_2}} \right)}{\omega_A \left(\omega_0 + \frac{S_1}{K_{D,A}^{S_1}} + \frac{S_2}{K_{D,A}^{S_2}} \right) + \omega_0 + \frac{S_1}{K_{D,I}^{S_1}} + \frac{S_2}{K_{D,I}^{S_2}}}. \quad (2)$$

Uit vergelijking 2 volgt dat S_1 het eiwit V activeert als $K_{D,A}^{S_1} < K_{D,I}^{S_1}$ (en hetzelfde geldt voor S_2). Het totaal aantal kopieën V van eiwit V dat in de actieve conformatie is, is dan $p_V^A V_T$.

Decodering

De decodering wordt gemodelleerd met 'Hill'-functies. Deze functies beschrijven de regulatie van twee (of meer) eiwitten, bijvoorbeeld de activatie van eiwit X_1 door eiwit V

$$X_1 = X_T \frac{V^n}{K^n + V^n}. \quad (3)$$

De belangrijke parameter is de drempelwaarde K die het transitiepunt bepaalt waarop de regulatie van V op X_1 verandert van minimaal tot maximaal (figuur 3d en e). Verder bepaalt n de breedte van dit regime; een grotere n leidt tot een smaller transitierégime. De variabele X_T bepaalt het maximale aantal kopieën. Het decodeerproces in ons model wordt beschreven door twee vergelijkingen:

$$X_1 = X_T \frac{V^2}{V^2 + K_1^2} \quad (4)$$

$$X_2 = X_T \frac{V^2}{V^2 + K_V^2} \frac{K_{X_1}^2}{X_1^2 + K_{X_1}^2} \quad (5)$$

Vergelijking 5 bestaat uit een vermenigvuldiging van de activatiefactor (door V) en de repressiefactor (door X_1).

32

mogelijk. Maar als we ook een additionele negatieve interactie tussen X_1 en X_2 (figuur 3a) toestaan, kan voor realistische parameters mapping C correct worden gedecodeerd.

In figuur 3c-e tonen we het decodeerprincipe. In deze figuur is K de drempelwaarde voor de concentratie waarbij activatie of juist repressie optreedt. X_1 wordt geactiveerd door V met drempelwaarde K_1 (vergelijking 4). De toestanden in X_1 worden door deze simpele activatie zo gegroepeerd dat de twee laagste toestanden van X_1 corresponderen met s_1^0 , en de twee hoogste met s_1^1 .

X_2 stijgt door activatie door V, met drempelwaarde K_V (vergelijking 5), maar, in tegenstelling tot X_1 , daalt X_2 door repressie door X_1 . Deze niet-monotone respons van X_2 als functie van V leidt tot een groepering van $x_2^0, x_2^1, x_2^2, x_2^3$ die aan de multiplexingvoorwaarde voldoet (figuur 1c). Inderdaad zijn de toestanden die

horen bij s_2^0, x_2^0 en x_2^1 , nu de twee laagste toestanden, en de toestanden x_2^2, x_2^3 , corresponderend met s_2^1 , zijn de hoogste.

Nu kunnen we ook begrijpen waarom het niet mogelijk is om de mappings A en B te decoderen. Voor mapping C is een responscurve voor X_2 als functie van V nodig waarbij de afgeleide één keer van teken verandert, terwijl voor mapping B en C de afgeleide minimaal twee keer van teken moet veranderen. Met slechts één incoherente feed-forwardlus is dit moeilijk te bewerkstelligen.

Decodering

Het is dus mogelijk om mapping C te decoderen, zodanig dat de twee laagste concentraties van X_1 (x_1^0, x_1^1) corresponderen met de afwezigheid van S_1 , de hoogste concentraties van X_1 (x_1^2, x_1^3) corresponderen met de aanwezigheid van S_1 , en soortgelijk voor X_2 .

De vraag is nu hoe mapping C geïm-

plementeerd wordt. Wij beschouwen het scenario waarin component V een receptoreiwit is dat aan de buitenkant van de cel de liganden S_1 en S_2 kan binden. Dit receptoreiwit schakelt tussen een actieve en een inactieve conformatie. Alleen in de actieve conformatie kan V X_1 en X_2 activeren. De toestand v^i in ons model correspondeert dan met de kans dat V actief is; dit is evenredig met het aantal receptormoleculen dat actief is. De kans dat één receptor 'actief' is, wordt vaak beschreven met behulp van een Monod-Wyman-Changeuxmodel (zie kader Mathematisch model).

De essentie van mapping C is het 'verwisselen' van de toestanden v^2 en v^3 . Normaliter verwacht je dat v^3 , waarbij beide signalen aanwezig zijn, het grootst is. In mapping C is echter v^2 , waarbij slechts signaal S_1 aanwezig is, groter dan v^3 . Deze verwisseling ontstaat door de volgende combinatie van ingrediënten. S_1 en S_2 kunnen

allebei aan V binden, maar niet tegelijkertijd – ze concurreren dus met elkaar. Daarnaast activeert S_1 , wanneer het aan de receptor gebonden is, de receptor meer dan S_2 , maar heeft S_2 een hogere bindingsaffiniteit voor de receptor. Dit betekent dat S_2 relatief vaak aan de receptor gebonden zal zijn, wanneer beide signalen aanwezig zijn. Hierdoor zal de activiteit van de receptor minder zijn dan wanneer S_1 alleen aanwezig is, maar meer dan wanneer alleen S_2 aanwezig is (figuur 3b).

De invloed van ruis

Tot dusverre hebben we gekeken naar een deterministisch netwerk waarin biochemische ruis afwezig is. Echter, doordat biochemische reacties en fysische interacties kansprocessen zijn, zijn biochemische netwerken intrinsiek ruisig. Daarom moeten we bepalen of het netwerk ook onder invloed van ruis betrouwbaar signalen kan multiplexen. De ruis resulteert in variaties in de concentraties van V , X_1 en X_2 . Over een lange tijdperiode komt de gemiddelde concentratie overeen met het ruisloze netwerk, maar op elk willekeurig tijdstip schommelt de concentratie rondom dit gemiddelde. Deze variatie beschrijven we met de standaarddeviaties σ_v , σ_{x_1} en σ_{x_2} . Een goede maat voor de betrouwbaarheid van het netwerk is de gedeelde informatie $I(S;X)$ tussen signaal S en respons X (zie kader *Gedeelde informatie*). De gedeelde informatie I beschrijft hoe betrouwbaar we de toestand van S kunnen bepalen, met kennis van de toestand van X . Als S twee toestanden heeft is de maximale $I(S;X)$ precies 1 bit. In het model betekent dit dus dat als de gedeelde informatie $I(S;X)$ 1 bit is, kennis van de toestand van X het mogelijk maakt de aan- dan wel afwezigheid van S met 100% zekerheid te bepalen.

Om de betrouwbaarheid te bestuderen, optimaliseren we met behulp van simulaties de gedeelde informatie $I(S;X_1)$ en $I(S_2;X_2)$ door de netwerkparameters ($\omega_A, K_{D,A}^{S_1}, K_{D,I}^{S_1}, K_{D,A}^{S_2}, K_{D,I}^{S_2}, K_1, K_V, K_{X_1}$, zie vergelijkingen 2, 4 en 5) te variëren.

Om $I(S;X)$ te berekenen zijn de conditionele distributies $p(X_1|S_1)$, $p(X_2|S_2)$ nodig (zie kader *Gedeelde informatie*). Een snelle methode om deze te berekenen is de aanname dat de conditionele distributies normaal

Gedeelde informatie

De entropie van een willekeurige variabele beschrijft de onzekerheid in die variabele [7]

$$H(X) = - \sum_x p(x) \log_2 p(x) \text{ [bits]}, \quad (6)$$

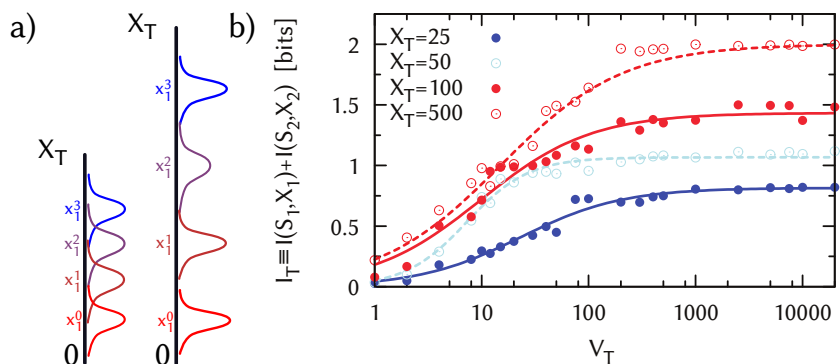
waar $p(x)$ de kansverdeling van X is. De entropie wordt gegeven in bits. Een variabele X met bijvoorbeeld 4 toestanden, ieder met kans $p=1/4$, heeft een entropie (onzekerheid) van 2 bits. Dit betekent dat je met 2 (slim gekozen) ja/nee-vragen altijd 100% zekerheid krijgt over de toestand van X . De conditionele entropie beschrijft de onzekerheid in de variabele X , gegeven de waarde van de variabele Y , met $p(x|y)$ de conditionele kansverdeling,

$$H(X|Y) = - \sum_y p(y) \sum_x p(x|y) \log_2 p(x|y). \quad (7)$$

De gedeelde informatie $I(X;Y)$ is het verschil tussen de onzekerheid in X ($H(X)$) en de conditionele onzekerheid $H(X|Y)$,

$$I(X;Y) = H(X) - H(X|Y) = \sum_{x,y} p(x,y) \log_2 \frac{p(x,y)}{p(x)p(y)}. \quad (8)$$

Als $H(X|Y) = 0$, met andere woorden als er tussen Y en X een een-op-een-relatie is, dan is de gedeelde informatie $I(X;Y)$ gelijk aan $H(X)$; kennis van Y geeft ons maximale informatie over X .



Figuur 4 a) Door de intrinsieke ruis wordt een gegeven ingangspatroon niet gemapped op één waarde van X_1 , maar op een verdeling van waarden van X_1 (en soortgelijk voor X_2 , niet getoond). Voor een groter totaal aantal kopieën X_i is de kans op overlap van de toestanden kleiner, omdat de standaarddeviatie van het kopieeaantal schaal met de wortel van het gemiddelde aantal. b) De som van de gedeelde informatie tussen S_1 en X_1 en tussen S_2 en X_2 als functie van het aantal kopieën in V , V_T , en X_1, X_2, X_T .

verdeeld zijn, zodat de verdeling volledig bepaald wordt door het gemiddelde en de standaarddeviatie. Een belangrijk argument ten gunste van deze aanname is dat deze leidt tot een ondergrens voor de gedeelde informatie. De conditionele distributies worden gekenmerkt door de gemiddelden v^j , x_1^j en x_2^j , en de standaarddeviaties σ_v^j , $\sigma_{x_1}^j$ en $\sigma_{x_2}^j$ (met $j=0, 1, 2, 3$). Echter, omdat het netwerk niet-lineair is (vergelijkingen 2, 4 en 5), is het niet mogelijk een analytische oplossing te bepalen voor de σ 's. Het is wél mogelijk om de ruistermen analytisch te bepalen voor een gelineariseerd

netwerk. Deze benadering, de Linear Noise Approximation [4], is snel en verbazingwekkend accuraat [5].

Een belangrijke voorwaarde voor betrouwbaar multiplexen is dat in de mapping van de ingangssignalen op de verschillende toestanden van V , X_1 en X_2 , deze laatste toestanden voldoende gescheiden blijven. Ruis in de reacties kan er echter toe leiden dat deze toestanden gaan overlappen (zie figuur 4a), wat de betrouwbaarheid van het netwerk kan verkleinen. Hoe lager het aantal kopieën van een eiwit, hoe groter de relatieve invloed van de ruis. We optimaliseren daarom het

netwerk voor verschillende aantallen V_T en X_T ($X_T = X_{1,T} = X_{2,T}$). Figuur 4b toont het resultaat. In de horizontale richting is de invloed van het aantal V_T -kopieën duidelijk zichtbaar. Als het totale aantal kopieën V_T te klein is, is het niet mogelijk om maximale betrouwbaarheid te verkrijgen. Eenzelfde effect zien we bij het verkleinen van het aantal kopieën X_T . In een typische biologische cel zijn normaliter al snel zo'n duizend kopieën van een eiwit aanwezig. Onze resultaten tonen dus aan dat onder biologisch relevante condities, cellen minimaal twee binaire biochemische signalen

betrouwbaar kunnen multiplexen. In dit artikel hebben we een multiplexingsschema besproken waarin de ingangssignalen gecodeerd worden in het concentratieniveau van een component van het signaaltransductienetwerk. Extra- en intracellulaire signalen worden echter niet alleen gecodeerd in concentraties van componenten, maar ook in frequenties van oscillaties of in de duur van een puls. Ook deze coderingsschema's kunnen gebruikt worden om verschillende signalen te multiplexen. Het is nu aan de experimentatoren om te bepalen of cellen ook daadwerkelijk deze multi-

plexingsschema's gebruiken, en zo ja, hoe betrouwbaar ze zijn.

Referenties

- 1 C. J. Marshall, *Cell* **80** (1995) 179.
- 2 E. Batchelor, A. Loewer, C. Mock en G. Lahav, *Molecular systems biology* **7** (2011) 488.
- 3 W. H. de Ronde, *Multiplexing Biochemical Signals* (VU University, Amsterdam, 2012).
- 4 N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (Elsevier, Amsterdam, 2001).
- 5 E. Ziv, I. Nemenman en C. Wiggins, *PLoS one* **2**, e1077 (2007).
- 6 J. Monod, J. Weyman en J.-P. Changeux, *Molecular Biology* **12** (1975) 88-118.
- 7 C. E. Shannon, *Bell System Tech. J.* **27** (1948) 379.