

Efektywność transmisji informacji w sieciach neuronowych w ujęciu teorii Shannona

Bartosz Paprocki

Instytut Mechaniki i Informatyki Stosowanej,
Uniwersytet Kazimierza Wielkiego

Janusz Szczepanski

Instytut Podstawowych Problemów Techniki,
Polska Akademia Nauk

Informacja (w sensie Shannona)

Wyprowadzona w oparciu o rozsądne/zapostulwane aksjomaty

Miara ilości **informacji** przenoszonej przez dane zdarzenie jest wielkość niepewności (co do wyników eksperymentu lub na przykład zrealizowanego sygnału wejściowego) usuniętej po zajściu tego zdarzenia.

Definicja

(MIARA NIEPEWNOŚCI)

Ilość informacji I otrzymanej przy zajściu zdarzenia x oblicza się jako:

$$I(x) = -\log p(x),$$

gdzie $p(x)$ jest prawdopodobieństwem zajścia zdarzenia x .

W takiej definicji znaczenie komunikatu (semantyka) jest pomijane.

Schemat przesyłania/przetwarzania informacji (kanał transmisyjny)

Krok 1: Źródło

ex.: Wiadomość, bodziec, ciąg znaków, obraz, **Kodowanie**

Krok 2: Kanał (zazumiony)

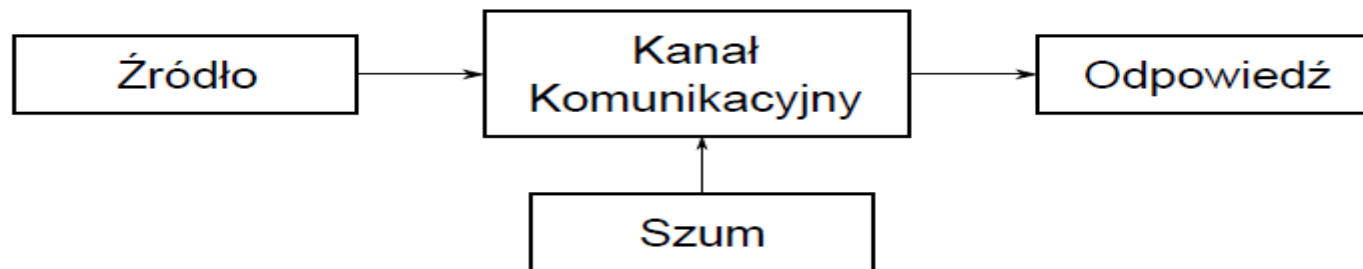
ex.: Przewód, komórka nerwowa, sieć neuronów

Krok 3: Odpowiedz - **Dekodowanie**

ex.: Otrzymana wiadomość, usłyszane słowo

Proces stochastyczny

Proces stochastyczny



Matematyczny opis kanału: Prawdopodobieństwa warunkowe:
wejścia-wyjścia (zależą od struktury kanału)

Miara efektywności Informacja Wzajemna

Entropia jest przeciętna (najbardziej prawdopodobna – oczekiwana) wartością informacji wzajemnie wykluczających się zdarzeń z pewnego zbioru.

Definicja

Entropia $H(X)$ zbioru X wykluczających się zdarzeń nazywamy Wielkość (wprowadzana aksjomatycznie):

$$H(X) = E(I(X)) = - \sum_{x \in X} p(x) \log p(x)$$

gdzie $E(I(X))$ jest wartością oczekiwaną informacji pochodzących ze zdarzeń $x \in X$, a $p(x)$ jest prawdopodobieństwem zajścia zdarzenia x .

I

$$0 \leq H(X) \leq \log |X|$$

Entropia warunkowa

- ▶ **Entropia warunkowa** mierzy entropię (niepewność) zbioru zdarzeń X pozostałą po zajściu zdarzenia ze zbioru Y .

Definicja

Entropią $H(X|Y)$ zbioru X pod warunkiem (zajścia zdarzenia ze zbioru) Y , jest wielkość:

$$H(X|Y) = \sum_{y \in Y} p(y)H(X|y) = - \sum_{y \in Y} p(y) \sum_{x \in X} p(x|y) \log p(x|y)$$

gdzie $p(x|y)$ jest prawdopodobieństwem zajścia zdarzenia x ze zbioru X , pod warunkiem zajścia zdarzenia y ze zbioru Y .

- ▶ $H(X|Y) = H(X, Y) - H(Y)$, gdzie $H(X, Y)$ jest entropią łączną zbiorów X i Y .

Redukcja niepewności co do sygnału na wejściu po informacji co było na wyjściu

Informacja wzajemna

- ▶ **Informacją wzajemną** dwóch zbiorów zdarzeń nazywamy ilość informacji o zdarzeniach z jednego zbioru zawartą w zdarzeniach z drugiego zbioru.

Definicja

Informacją wzajemną $I(X; Z)$ zbioru zdarzeń Z o zbiorze zdarzeń X , jest wielkość:

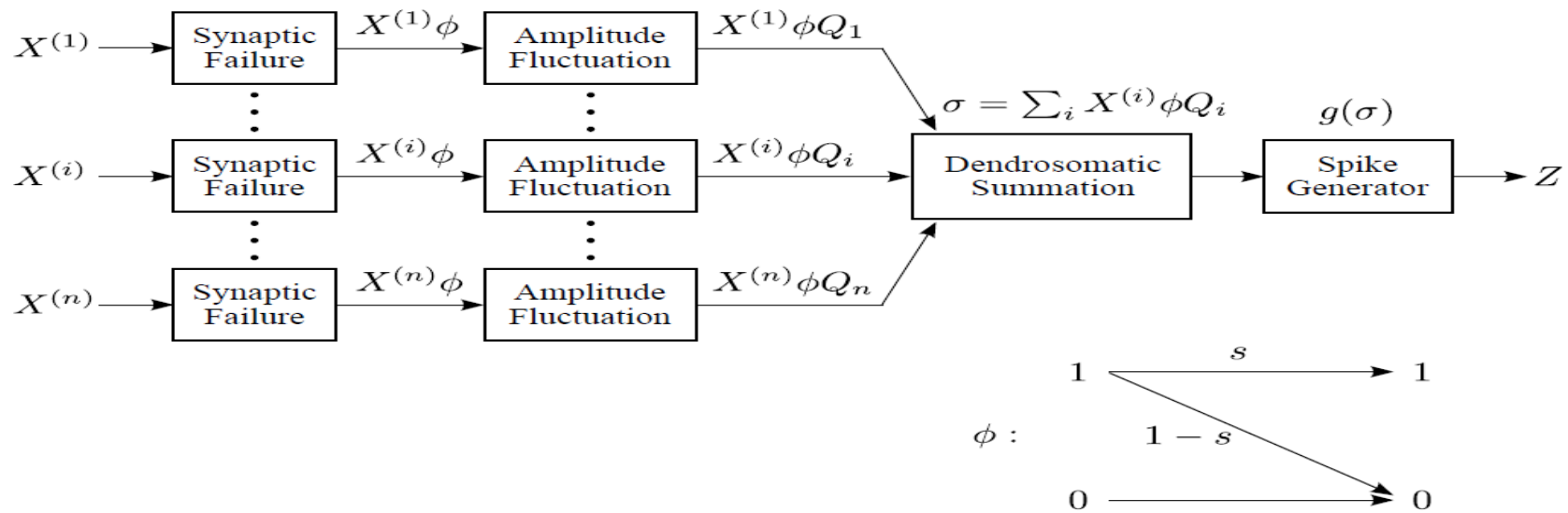
$$I(X; Z) = H(X) - H(X|Z) = H(X) + H(Z) - H(X, Z),$$

gdzie $H(X)$ jest entropią X , $H(X|Z)$ entropią warunkową X pod warunkiem Z , a $H(X, Z)$ entropią łączną X i Z .

Do estymowania entropii wybrany został Estymator Stronga (PRL, 1998)

Levy, Baxter : Journal of Neuroscience 2002

Model komórki nerwowej



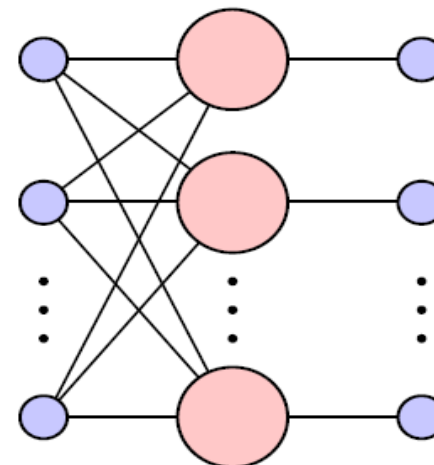
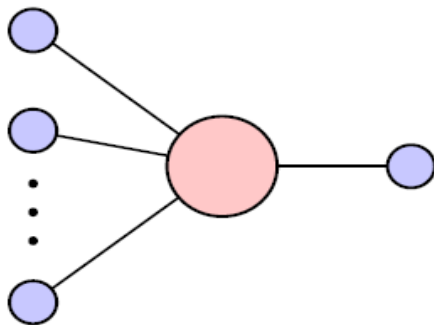
Parametry: **s**-synaptic failure, **g**-threshold, **Qi**-Amplitude fluctuations

PROSTE SIECI

W oparciu o ten model rozpatrywano również sieci takich komórek.

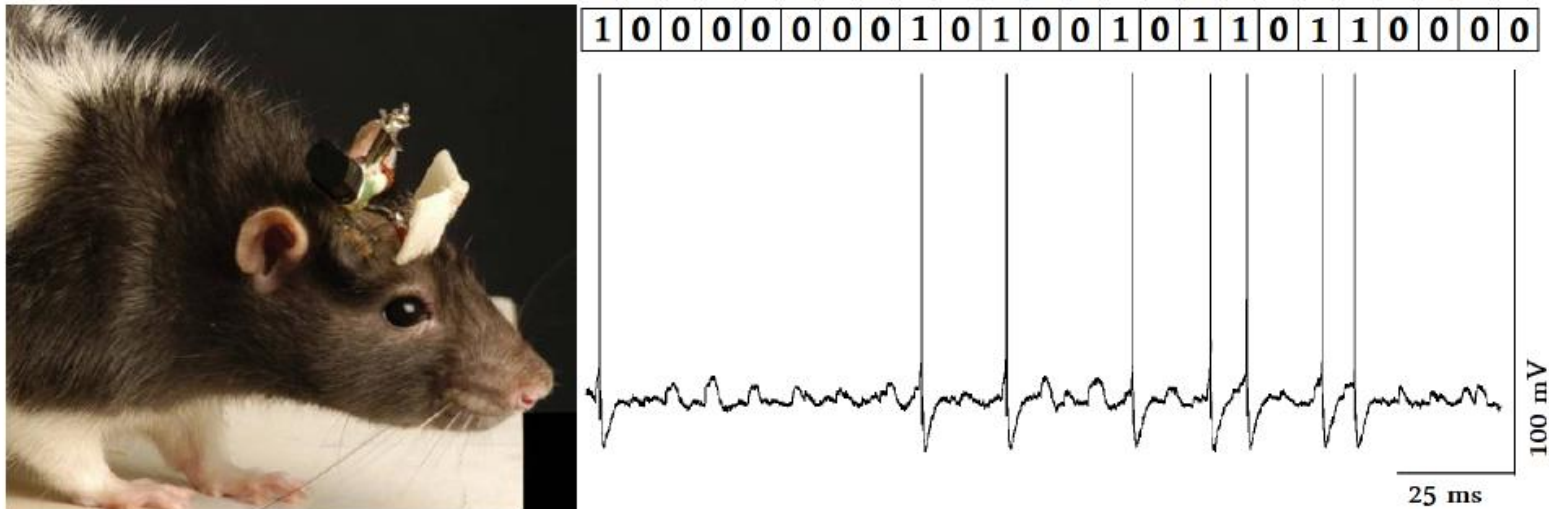
Sieci jednowarstwowe

- ▶ **X** staje się wejściem dla większej ilości neuronów.
- ▶ Każdy neuron oblicza swoje pobudzenie niezależnie od innych.
- ▶ W efekcie na wyjściu otrzymuje się nie pojedyncze Z , ale $\mathbf{Z} = [Z^{(1)}, \dots, Z^{(m)}]$, gdzie m jest liczbą komórek w sieci.



Dwa modele źródeł

- ▶ Zdecydowano się skorzystać z dwóch modeli źródeł: zależne i niezależne.
- ▶ W obu przypadkach ograniczono alfabet do dwóch znaków:
 - 1: aktywacja progowa (ang. *spike*)
 - 0: brak aktywacji (*no-spike*)



Cel badań

Celem badań jest wyznaczenie optymalnych parametrów komórek i sieci neuronowych, analizując wartości **informacji wzajemnej** pomiędzy wejściem (bodźcami) a odpowiedzią struktury na wyjściu (pobudzeniem).

Estymowana informacja wzajemna

- ▶ Entropia warunkowa ze wzoru: $I(X; Z) = H(X) - H(X|Z)$ jest trudna do wyznaczenia.
- ▶ Z własności entropii: $H(X|Z) = H(X, Z) - H(Z)$, gdzie $H(X, Z)$ jest entropią łączną zmiennych X i Z .
- ▶ Wyznaczanie entropii łącznej jest zadaniem prostszym. Istnieją efektywniejsze metody (estymatory).
- ▶ Otrzymujemy: $I(X; Z) = H(X) + H(Z) - H(X, Z)$.

Tempo entropii

Niech X będzie pewnym procesem stochastycznym. Wtedy:

$$H(X) = \lim_{n \rightarrow \infty} \frac{1}{n} H(X_1, X_2, \dots, X_n).$$

Równoważnie:

$$H'(X) = \lim_{n \rightarrow \infty} H(X_n | X_{n-1}, X_{n-2}, \dots, X_1).$$

- ▶ Jeśli proces jest stacjonarny, wtedy: $H(X) = H'(X)$.

Problemy obliczeniowe – ogólnie

- ▶ Obliczenia są bardzo czasochłonne i wymagają sporej pamięci operacyjnej.
- ▶ Symulacje odbywały się na serwerze: 16 procesorów o mocy 2,67 GHz, 64 GB RAM.
- ▶ Wygenerowanie danych do jednego wykresu $I(f_r, s, g)$ wymaga od kilku do kilkudziesięciu godzin.

Duża dokładność modeli numerycznych dla eksperymentów raczej trudno osiągalna

SYNAPTIC FAILURE EFFECTS (SZUM) Niemontoniczność względem błędu w synapsach

Na osiach są: **fr** =firing rate, **s**=synaptic failure, **Poziomnice MI**

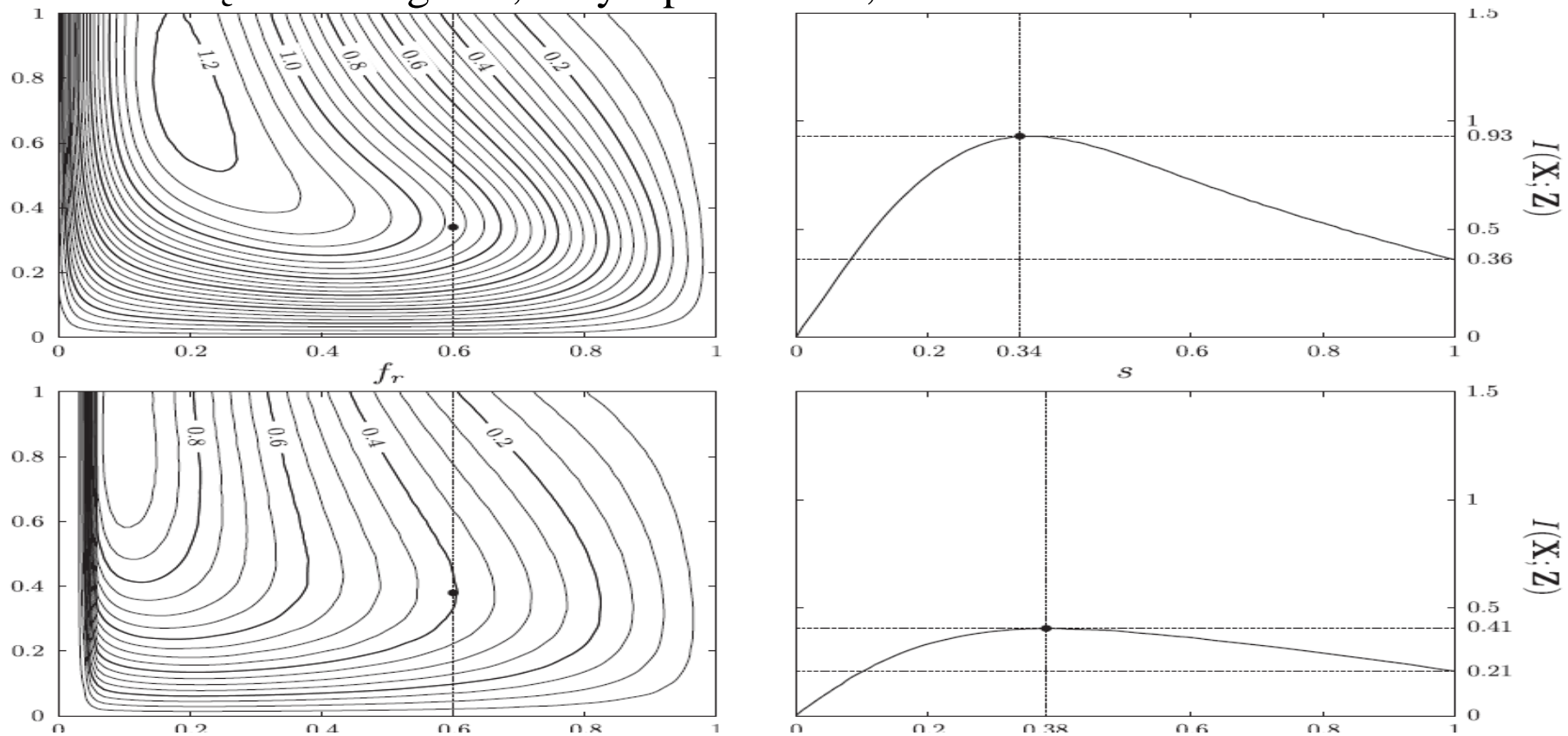


Fig. 4. Mutual Information dependencies on *synaptic success*, s , in neural network 5×5 architectures with threshold set to 5%. All plots (left column) are sliced at $f_r = 0.6$ and in all three cases Mutual Information values are maximized for *synaptic successes* less than 0.4, where they reach values about twice as large as for $s = 1$. Bernoulli source (top), Markov source with $p_{01} = 0.05$ (middle) and Markov source with $p_{01} = 0.10$ (bottom).

Ciągła linia – MI bez synaptic failure ($s=1$), *Kropki* – MI z synaptic failure

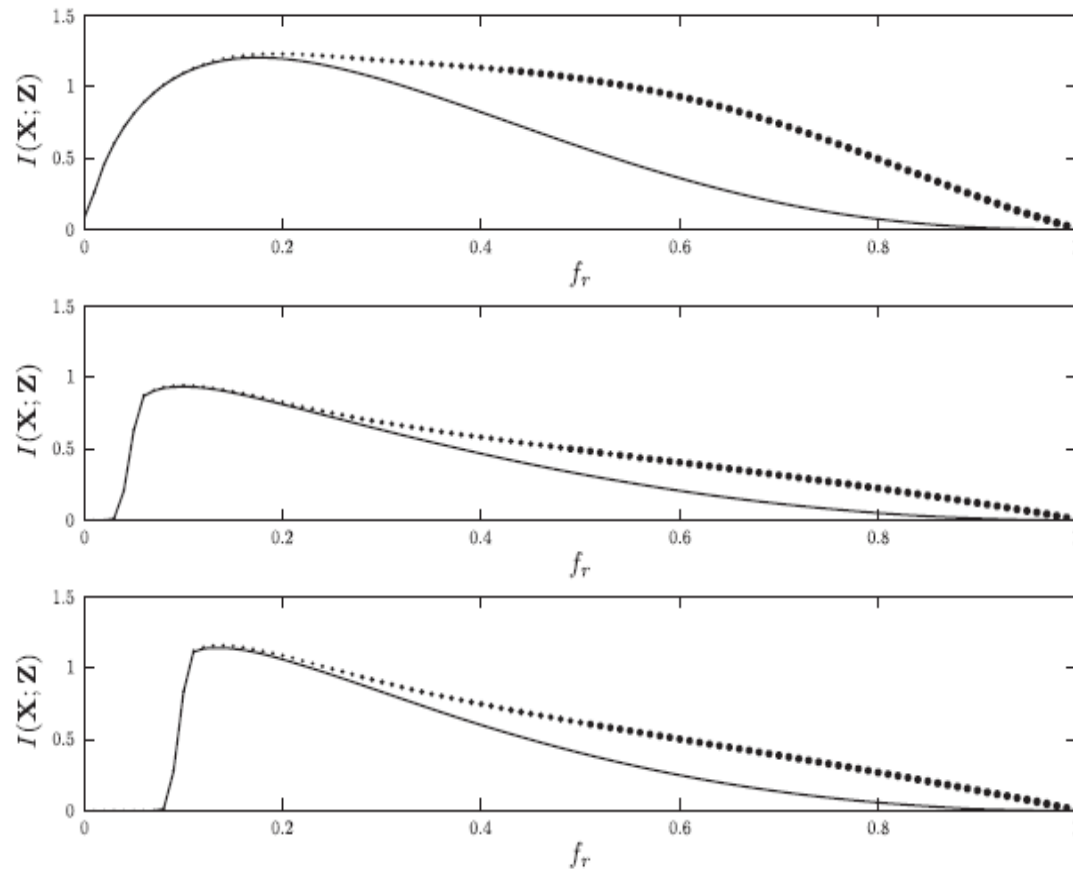


Fig. 5. Mutual Information dependencies on synaptic success, s , in neural network 5×5 architectures with threshold set to 5%. Comparison of maximal Mutual Information values (dotted line) with these achieved at $s=1$ (solid). Size of a dot is proportional to $1-s$, indicating the bigger the dot, the corresponding Mutual Information value is achieved at lower s . Bernoulli source (top), Markov source with $p_{01} = 0.05$ (middle) and Markov source with $p_{01} = 0.10$ (bottom).

THRESHOD ACTIVATION EFFECTS

Niemonotoniczność MI (max) względem progu aktywacji neuronu

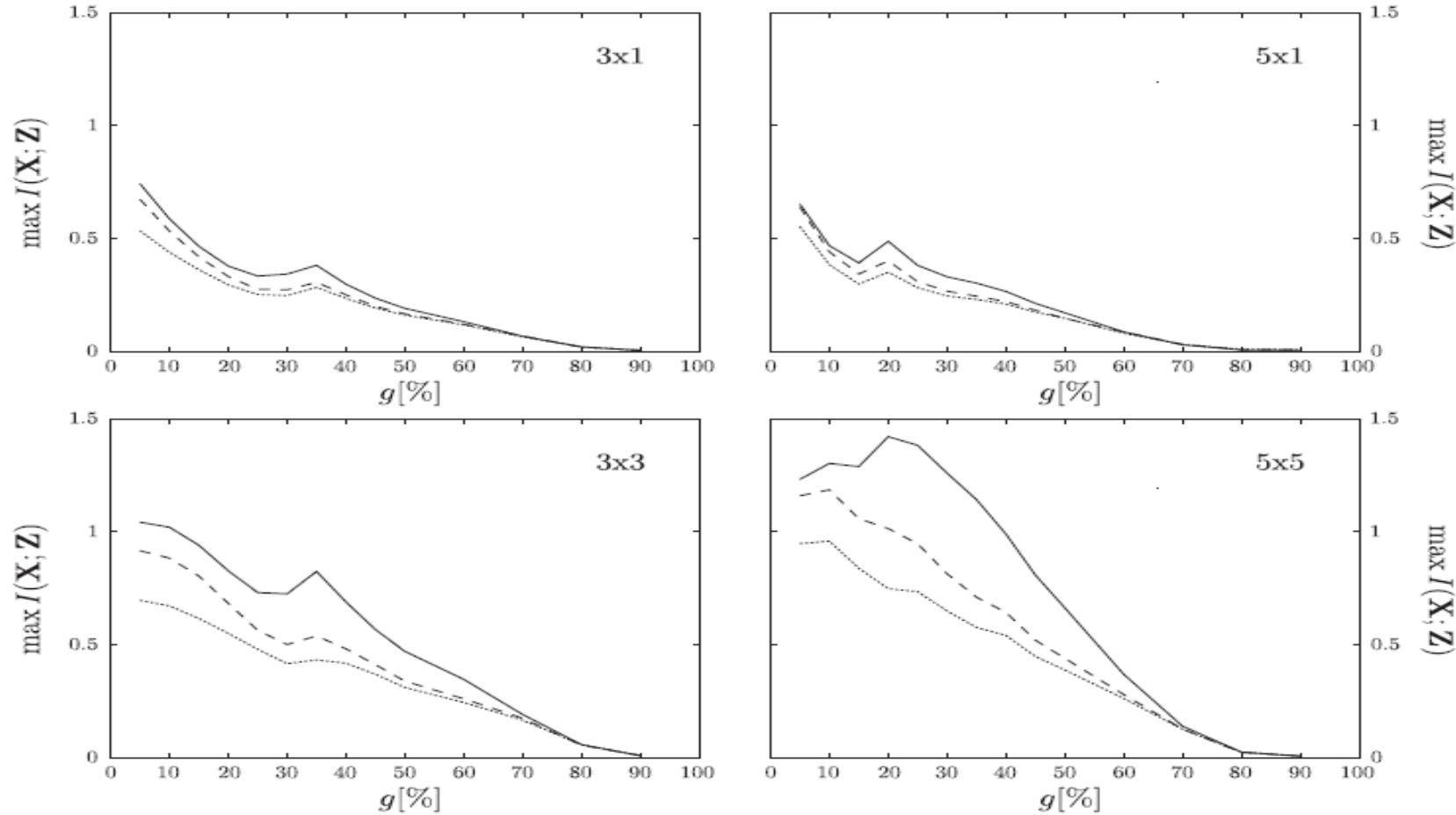


Fig. 7. Mutual Information's non-monotonic dependencies on *threshold* g , in all four discussed neural network architectures, indicating presence of local extrema. Position of these extrema seems to be independent of input stimuli. This is observed for all explored structures but the most complex one, the 5×5 network, where some of this regularity is lost, Bernoulli source (solid line), Markov with parameter $p_{01} = 0.05$ (dotted), Markov with $p_{01} = 0.10$ (dashed).

AMPLITUDE FLUCTUATIONS EFFECTS Qi

Funkcja gęstości Q_i

łumienie

wzmacnianie

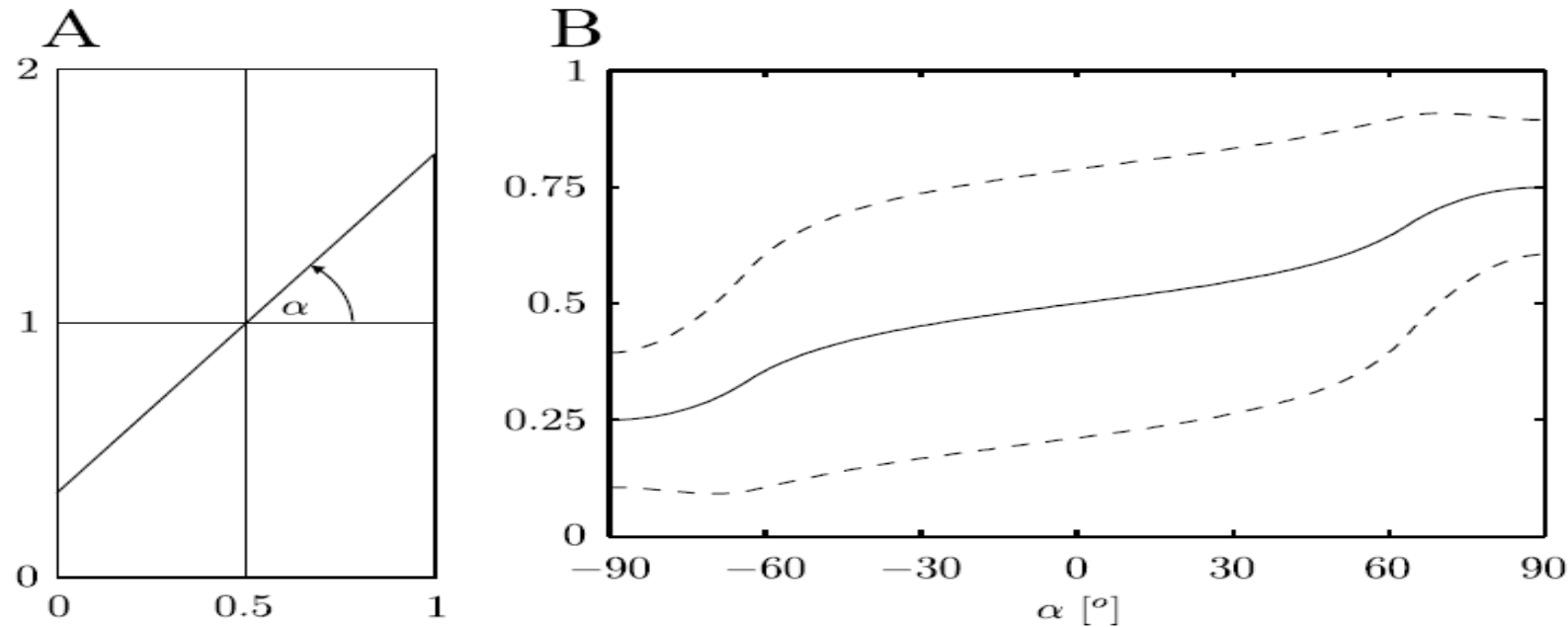


Fig. 3. Q_i and its characteristics, parameterised by the angle α . A, $\alpha \in [-90^\circ; 90^\circ]$ determines distribution of Q_i . For $\alpha = 0^\circ$, $Q_i(\alpha)$ is distributed uniformly on interval $[0; 1]$, this situation is covered in our previous paper [13]. Other interesting cases include especially: $\alpha = \pm 90^\circ$ and $\alpha = \arctan(\pm 2)$. B, basic probabilistic characteristics of $Q_i(\alpha)$: expected value (solid line) \pm standard deviation (dashed lines).

AMPLITUDE FLUCTUATIONS EFFECTS Qi cd

MI (max) w funkcji alpha, czyli od tłumienia do wzmacniania

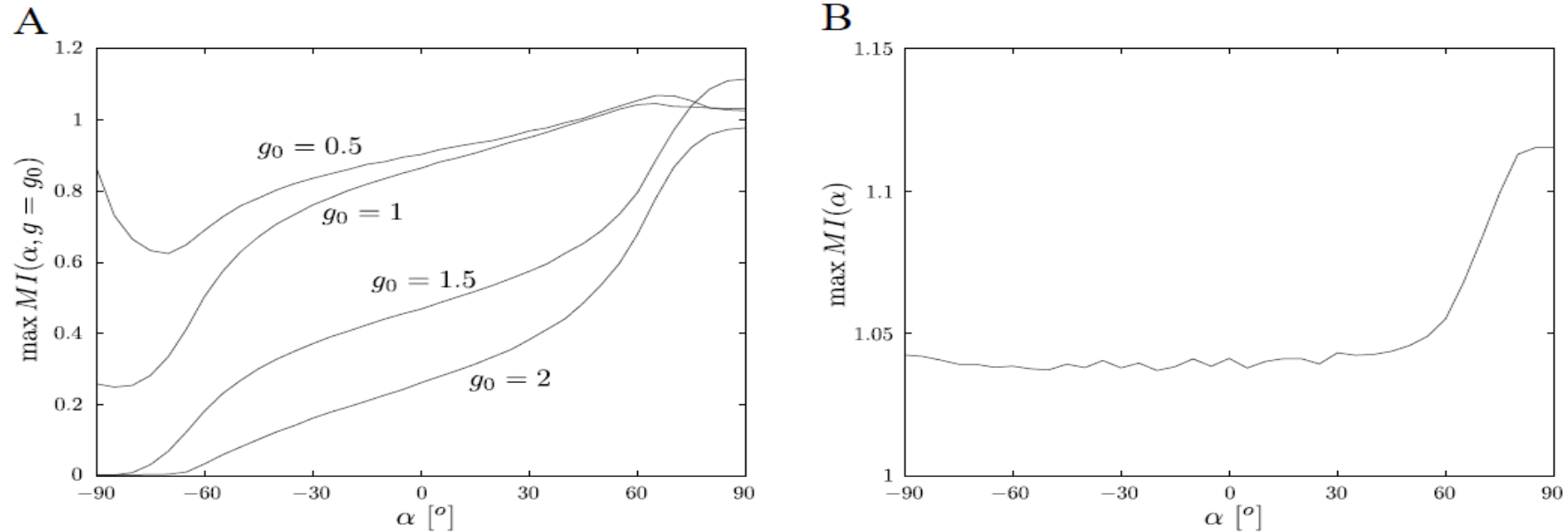
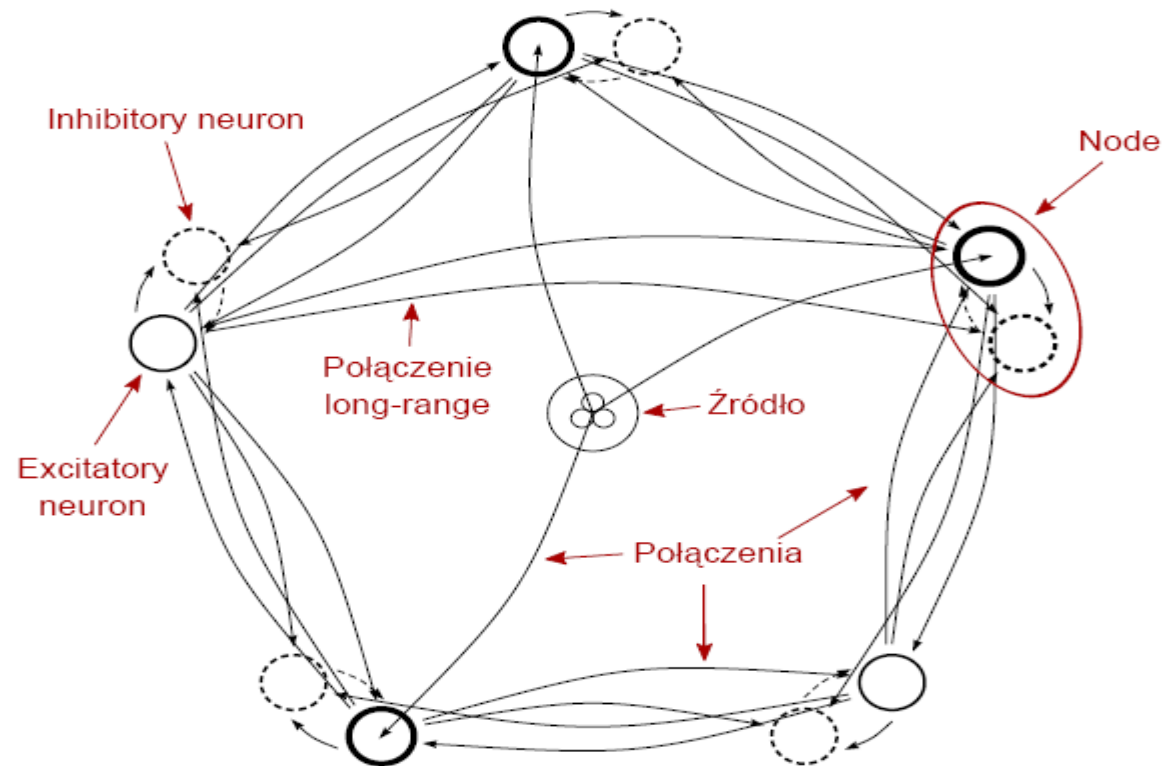


Fig. 4. A, Mutual Information in function of α maximized over f_T and s , for four different thresholds fixed at levels $g_0 = 0.5, 1, 1.5, 2$. B, Mutual Information in function of α maximized over all thresholds g , in addition to f_T and s . For angles $\alpha \in [-90^\circ, 45^\circ]$ it remains at level around 1.04 and raises to reach the value of ≈ 1.12 for angles $\geq 80^\circ$.

BRAIN LIKE MODELS

Schemat modelu (przykład)



PARAMETRY ARCHITEKTURY: **size/delay r** , **Excitatory-Inhibitory $b=I/E$**

Przyjęto:

OPÓŹNIENIE PROPORCJONALNE DO ODLEGŁOŚCI GEOMETRYCZNEJ

EXCITATORY-INHIBITORY Energy „COMMUNICATION”

Funkcja sumy (σ) przed progiem aktywacji neuronu ulega modyfikacji:

$$\sigma = \sum_{i=1}^n X^{(i)} \phi Q_i + \sum_{v=1}^w E^{(v)} \phi Q_v - b \cdot I \phi Q_I.$$

- ▶ Pierwszy składnik związany jest ze źródłem. Dotyczy tylko neuronów E . Jest on identyczny jak w Levy & Baxter (2002)
- ▶ Drugi składnik jest związany z bodźcami pochodzącymi od neuronów E , od których dana komórka odbiera sygnały (jest ich w)
- ▶ Trzeci składnik pochodzi od neuronu I (dot. wszystkich neuronów E)

Ciekawe problemy i pytania

- ▶ Znaczenie neuronów typu *inhibitory*.
- ▶ Wpływ opóźnień na efektywność transmisji. W szczególności rola wielkości skali. Znaczenie połączeń krótko- i długozasięgowych.
- ▶ Problem *Excitation-Inhibition balance*.
- ▶ Rozważania energetyczne. Zestawienie wydajności transmisji komórek/sieci z kosztami energetycznymi wytwarzania impulsów nerwowych by tę wydajność osiągnąć i utrzymać.
- ▶ Analiza optymalnych wartości parametrów opisujących: błędy w synapsach, próg aktywacji, komunikację wewnątrz węzłów.
- ▶ Zależność od ilości węzłów.

SIZE EFFECTS

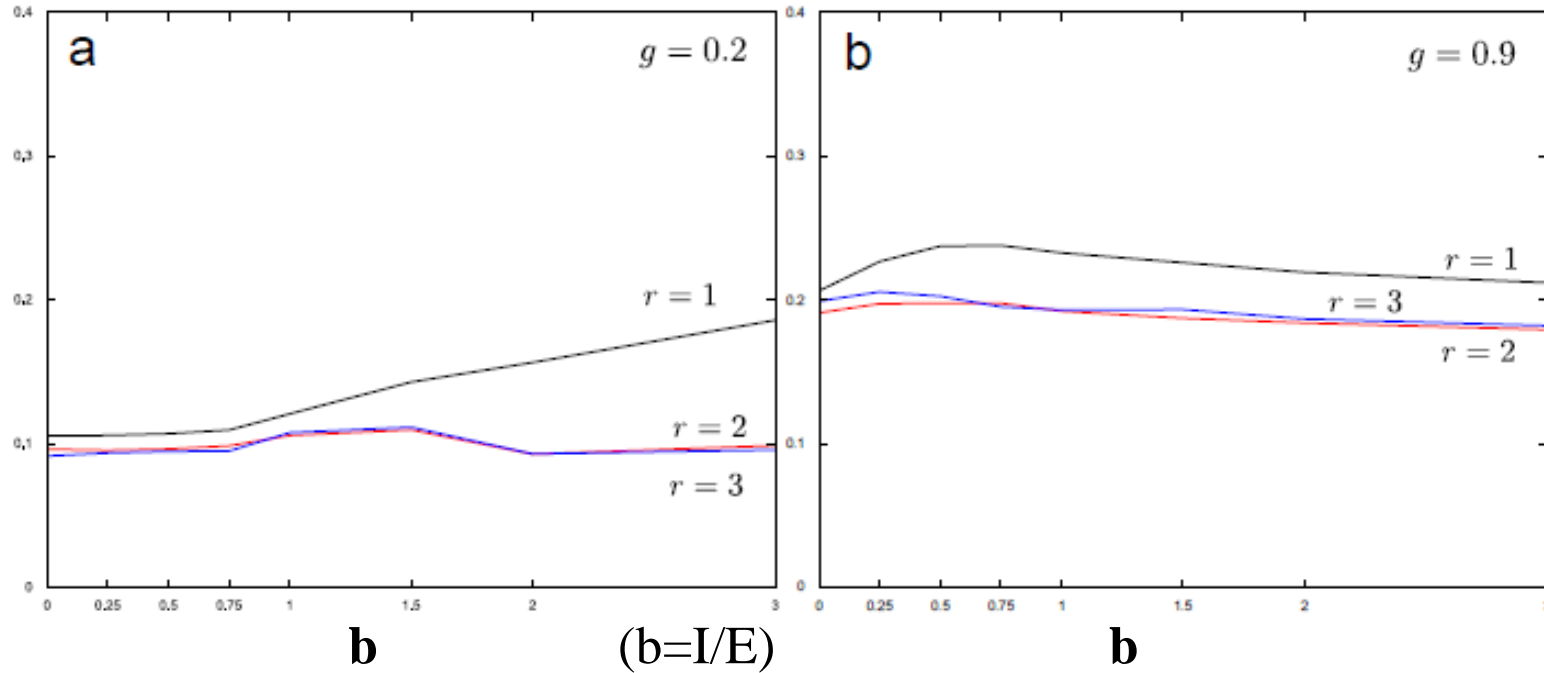


Figure 4: Zależność maksymalnej informacji wzajemnej od wartości tłumienia neuronu excitatory przez odpowiadający mu neuron inhibitory, wyrażonej przez stosunek sił oddziaływania neuronu I do E . Osie nie są opisane: na osi odciętych (poziomej) mamy $b = I/E$, na osi rzędnych (pion) mamy $\max MI$. Wykresy są dla neuronu E nr 1 (na godz. 12).

INFORMACYJNO-ENERGETYCZNA OPTYMALIZACJA

Following energy formulae were used for the simulations:

- For neurons E (excitatory) without access to the source X :

$$\frac{MI}{\vartheta} = \frac{MI}{s \cdot (b f_I + \sum_w f_w)}, \quad (1)$$

- Neurons E with access to the source X :

$$\frac{MI}{\vartheta} = \frac{MI}{s \cdot (n f_r + b f_I + \sum_w f_w)}, \quad (2)$$

- Neurons I (inhibitory):

$$\frac{MI}{\vartheta} = \frac{MI}{s \cdot \sum_w f_w}, \quad (3)$$

where:

- s denotes synaptic success,
- f_I – firing-rate of inhibitory neuron (from the same node as a given neuron E),
- f_w – firing-rate of w th excitatory neuron, predecesing a given neuron,
- f_r – firing-rate of the source X ,
- n – size of the source.

INHIBITORY EFFECTS IN INFORMATION/ENERGY EFFICIENCY

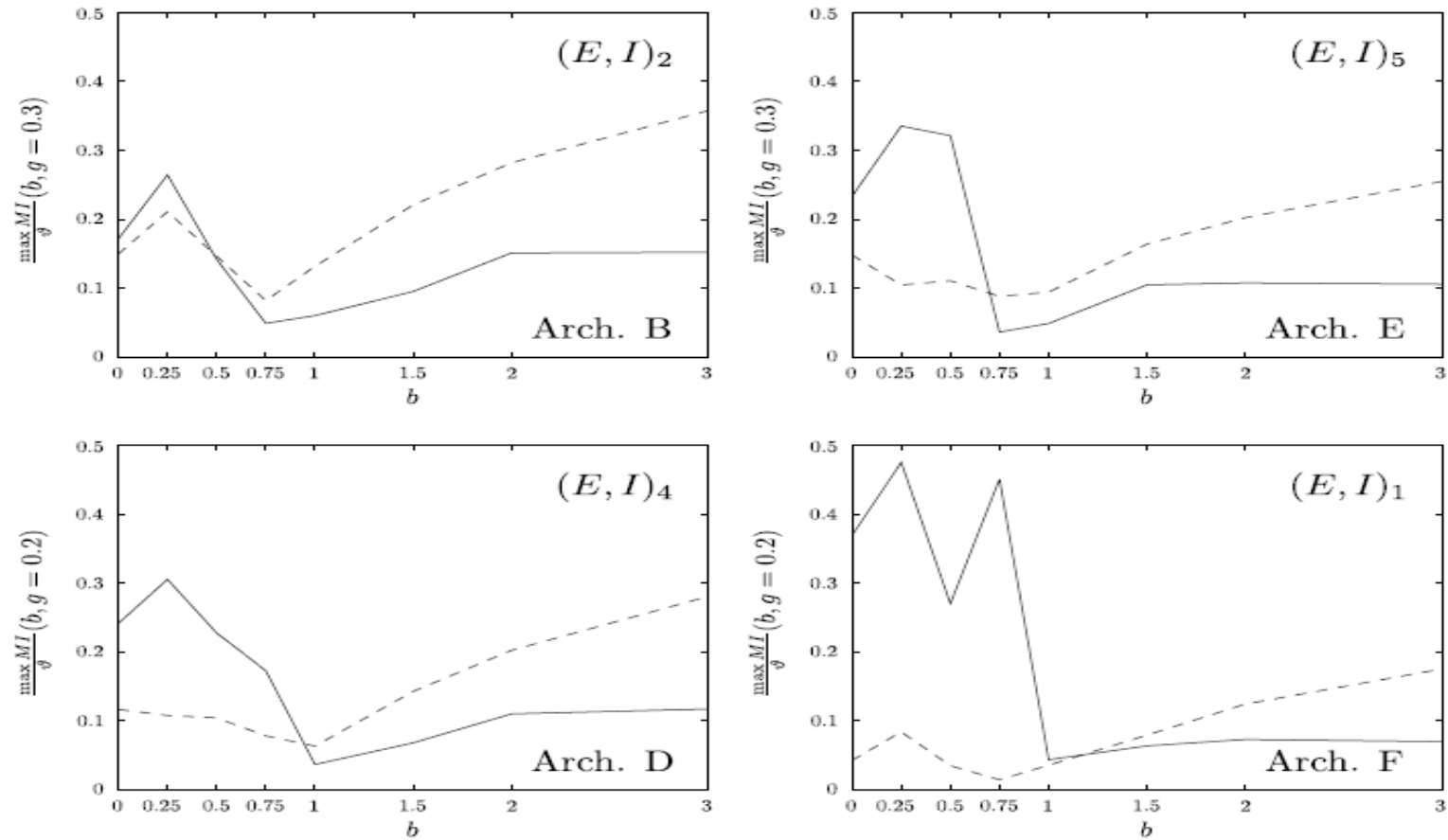


Figure 4: Examples of non-monotonicity of $\frac{\max MI}{\vartheta}(b, g)$ for 4 different architectures (clockwise, starting at top-left: B, E, F, D), each for one of structures' five nodes. Excitatory neurons are drawn with solid lines, inhibitory with dotted lines.

OBSERWACJE/WNIOSKI

Szum w synapsach może poprawiać efektywność transmisji nawet dość znacznie

Efektywność jest funkcją niemonotoniczną w funkcji progu – potwierdza, że informacja jest zawarta nie tylko w częstościach aktywacji

Inhibitory poprawiają istotnie efektywność transmisji (przy odpowiednich parametrach)

Istotne znaczenie skali – efekty opóźnień – w małych skalach znaczna poprawa efektywności, potem stabilizacja!

Wzmacnianie sygnałów (Q_i) niekoniecznie prowadzi do zwiększenia efektywności

KIERUNKI BADAŃ

- ▶ Przeprowadzenie badań dla bardziej złożonych architektur, modelujących strukturę mózgu. Sieci wieloneuronowe, wielowarstwowe. Skorelowane neurony, połączenia wewnątrzwarstwowe.
- ▶ Badania nad redundancją transmisji w komórkach i sieciach neuronowych.
- ▶ Analiza efektu głębokości korelacji źródła.