行政院國家科學委員會補助專題研究計畫 ■ **成 果 報 告**
□**期中進度報告**

在有限存取合作式多重輸入多重輸出環境下的非同調編碼理論與技術之研究
Noncoherent-based Coding Theory and Technique under Limited Cooperative MIMO Environment

計畫類別：□ 個別型計畫 ■ 整合型計畫
計畫編號：NSC 98－ 2221－ E－ 009－ 060－MY3
執行期間： 98 年 8 月 1 日至 101 年 7 月 31 日

計畫主持人：陳伯寧教授
共同主持人：王忠炫副教授
計畫參與人員： 施沛渝、王士瑋、吳庭伊、楊翊弘、黃柏元、陳品翰、楊俊彥、劉勁甫、呂志文

成果報告類型(依經費核定清單規定繳交)：□精簡報告 ■完整報告

本成果報告包括以下應繳交之附件：
□赴國外出差或研習心得報告一份
□赴大陸地區出差或研習心得報告一份
■出席國際學術會議心得報告及發表之論文各一份
□國際合作研究計畫國外研究報告書一份

處理方式：除產學合作研究計畫、提升產業技術及人才培育研究計畫、列管計畫及下列情形者外，得立即公開查詢
□涉及專利或其他智慧財產權，□一年■二年後可公開查詢

執行單位：國立交通大學電信工程研究所

中 華 民 國 101 年 7 月 31 日

1. 中、英文摘要及關鍵詞(keywords)

關鍵詞：多重輸入多重輸出系統；有限存取；通道容量；功率配置；注水式功率配置；非同調編碼

隨著無線通訊產業的蓬勃發展，穩定傳輸、高傳輸率以及高移動性三項特性兼具之通訊技術已成為用戶端對下世代無線傳輸技術之基本要求。因此，在本計畫的三年研究中我們分別從兩大方面著手探討適用於快速移動環境之最佳化通訊系統設計，以達到提供高通訊品質與高傳輸速率的目的。其一為結合通道估測與錯誤更正之時空碼編碼設計與其非同調解碼設計，其二為新式的有限存取多終端系統分析與傳輸設計，該系統可免除傳統通訊系統中高速移動載具漫遊於多個基地台所需的多次換手機制，進而提高傳輸速率與通訊品質。經由三年的研究，對於結合通道估測與錯誤更正碼的部分我們提出一個系統化的演算法能夠找出效能優異的非同調碼字；對於解碼部分我們也提供了一個循序解碼演算法能大幅降低所需之解碼複雜度。對於新式的有限存取多終端系統之研究，我們已確立了此系統在任意通道下之通道容量與相對應的最佳功率配置機制，而對於通道考量為"相加雜訊家族"時，其最佳功率配置可由"二階段注水式配置準則"而得，此配置準則可視為消息理論中註明的注水式功率配置之推廣與延伸。此外，藉由"二階段注水式配置準則"我們亦找出任意通道雜訊程度之定義並探討了該定義在總功率趨近於零與趨近於無限大時的分析。

Keywords: multiple-input multiple-output, limited-access, channel capacity, power allocation, water-filling, non-coherent codes

As the developments of wireless communication, stability in quality, high data rates and high mobility have become basic requirements in next generation communication technologies. In order to satisfy the three requirements simultaneously, we focus on two research topics in this three-years project. One is the design of combining channel estimation and error correcting space-time code, as well as its non-coherent decoder. The other highlights the analysis and design of a multiple-terminal system with limited access, which is a situation that may encounter in a highly mobile environment. In these three years, we have proposed a systematic algorithm to generate codewords instead of doing computer searching in the research of the first topic, and we also provided its maximum-likelihood-decoding algorithm with low complexity. For the analysis of the novel system with limited access, we have derived the channel capacity for general channel models and found the optimal power allocation to achieve the capacity. Moreover, when the channel model is reduced to additive noises of the same family, we found that the optimal power allocation can be obtained by a simple two-phase water-filling process. Finally, following the interpretation of two-phase water-filling, we can further characterize the degree of "noisiness" for general channels and analyze the degree of noisiness when total power is sufficiently small and large, respectively.

2. 計畫回顧

2.1 Reviews of the work in the first year:

In the first year, we compared our non-coherent code design under several scenarios with Xu's code, which is specifically designed for a frequency-nonselective OFDM system (while our systematic code construction scheme can also be applied in a frequency selective environment). Our simulation results indicate that a blind-detectable noncohrent code can really be made robust for channels whose taps vary more often than a coding block. A side advantage of our code construction scheme is that its systematic structure makes it maximum-likelihoodly decodable by the priority-first search algorithm. Thus, when being compared with the operation-intensive exhaustive decoder, the decoding complexity is greatly reduced especially when codes of longer code length is adopted.

2.1.1 The system model:

Suppose that a codeword $\boldsymbol{b} = [b_1 \ \cdots \ b_N]^T$ is transmitted over a block fading channel of memory order P, of which channel coefficients vary in every Q symbols, where $b_i \in \{\pm 1\}$ and $Q > P$. By letting $L \triangleq N + P$ and $M \triangleq \lceil L/Q \rceil$, the system can be modelled by:

$$\boldsymbol{y} = \mathbb{B}\boldsymbol{h} + \boldsymbol{n},$$

where $\boldsymbol{n}$ is zero-mean white Gaussian distributed, $\boldsymbol{h} \triangleq \left[\boldsymbol{h}_1^H \ \ \boldsymbol{h}_2^H \ \ \cdots \ \ \boldsymbol{h}_M^H\right]^H$ with $\boldsymbol{h}_k \triangleq [h_{0,k} \ \ h_{1,k} \ \ \cdots \ \ h_{P,k}]^T$, and

$$\mathbb{B} \triangleq \mathbb{B}_1 \oplus \mathbb{B}_2 \oplus \cdots \oplus \mathbb{B}_M$$

with $\mathbb{B}_k \triangleq [\mathbf{0}_{Q\times P} \ \ \mathbb{I}_Q][\boldsymbol{b}_k \ \ \tilde{\mathbb{E}}\boldsymbol{b}_k \ \ \cdots \ \ \tilde{\mathbb{E}}^P\boldsymbol{b}_k]$. Here, $\mathbf{0}_{Q\times P}$ represents a $Q \times P$ all-zero matrix, $\mathbb{I}_Q$ is a $Q \times Q$ identity matrix, $\boldsymbol{b}_k \triangleq [b_{(k-1)Q-P+1} \ \ \cdots \ \ b_{(k-1)Q+1} \ \ \cdots \ \ b_{kQ}]^T$ is a portion of the transmitted codeword $\boldsymbol{b}$,

$$\tilde{\mathbb{E}} \triangleq \begin{bmatrix} 0 & 0\cdots & 0 & 0 \\ 1 & 0\ddots & 0 & 0 \\ 0 & 1\ddots & 0 & 0 \\ 0 & 0\cdots & 1 & 0 \end{bmatrix}_{(Q+P)\times(Q+P)}$$

equates the logical left-shift operator, and "$\oplus$" is the direct sum operator of two matrices. Also, for notational convenience, we let $n_j = 0$ for $j > L$, and $b_j = 0$ for $j \le 0$ and $j > N$. Under such system setting, $\boldsymbol{y}$ is an $MQ \times 1$ received vector with $y_j = 0$ for $j > L$.

It can be derived that the joint maximum-likelihood decoder upon the reception of $\boldsymbol{y}$ is given by:

$$\hat{\boldsymbol{b}} = \arg\max_{\boldsymbol{b}\in\mathcal{C}} \sum_{k=1}^{M} \left\| \boldsymbol{y}_k \boldsymbol{y}_k^H - \boldsymbol{P}_{B_k} \right\|^2, \quad (1)$$

where $\boldsymbol{y}_k \triangleq [y_{(k-1)Q+1} \ \ y_{(k-1)Q+2} \ \cdots \ y_{kQ}]$ is the output portion affected by $\boldsymbol{b}_k$, and $\mathbb{P}_{B_k} \triangleq \mathbb{B}_k(\mathbb{B}_k^T\mathbb{B}_k)^{-1}\mathbb{B}_k^T$. In the above derivation, we assume that the receiver, although it knows nothing about $\boldsymbol{h}$, has perfect knowledge about the values (or the upper bounds) of P and Q.

2.1.2 Code construction:

Based on years of research efforts, we already have some knowledge in the construction of non-coherent codes for $P=0$ (frequency nonselective) and $P=1$ (frequency selective). For completeness of this report, we list the code generating algorithm below.

Step 1. Fix $b_1 = -1$,[1] and choose a certain integer Δ defined later. Find 2^K codewords of the (N, K) code by repeating Steps 2–4 for $0 \le i \le 2^K - 1$.

Step 2. Let $\rho_{\min} = 0$ and $\rho = i \cdot \Delta$.

Step 3. For $\ell = 2$ to N, assign the ℓ-th bit of the i-th codeword, b_ℓ, according to that if $\rho < \rho_{\min} + \gamma_\ell$, then $b_\ell = -1$; else, $b_\ell = 1$ and $\rho_{\min} = \rho_{\min} + \gamma_\ell$, where

$$\gamma_\ell = |\mathcal{A}_P(b_1, \ldots, b_{\ell-1}, b_\ell = -1)|,$$

which will be defined shortly.

Step 4. Store the ith codeword $\boldsymbol{b}$, and goto Step 2 for the next codeword until all 2^K codewords are selected.

In the above coding design, the Δ-th codeword must be of the form $[\underbrace{-1 \cdots -1}_{K+1} \ 1 \ 1 \ \boldsymbol{u} \ 1]$, where $\boldsymbol{u}$ is a maximum-length shift-register sequence. When our code is compared with the three-times-repetitive (12, 6) code proposed by Xu *et al*, we found that when the channel coefficients remain constant over the entire coding block, the proposed (36, 6) code performs 0.7 dB better than Xu's code as shown in Figure 1. More details can be found in [3].

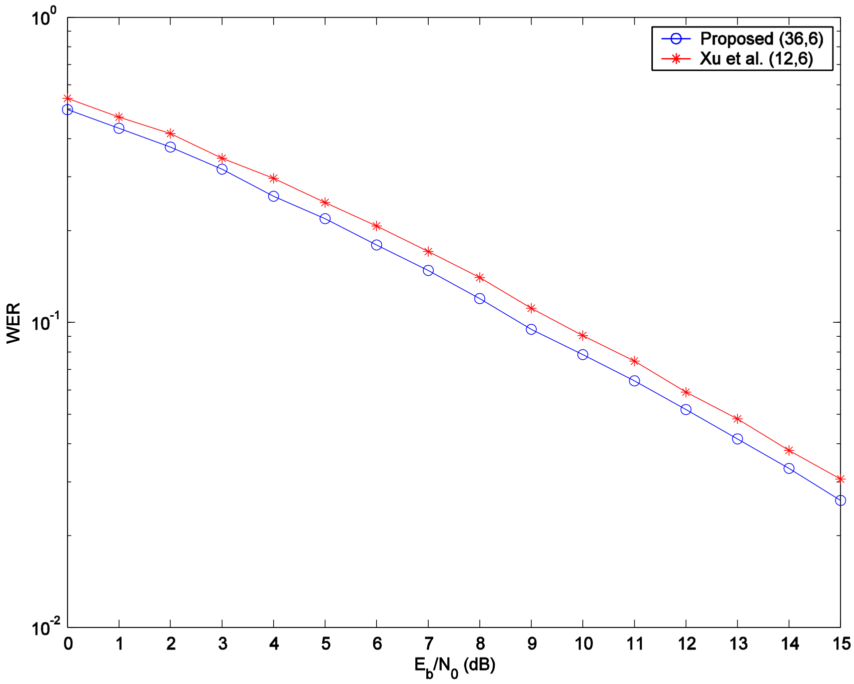

Figure 1: Word error rates (WERs) for the constructed (36, 6) code and the three-times-repetitive (12,6) code proposed by Xu *et al* over flat fading channel with channel coefficients unchanged during the transmission of a codeword.

2.1.3 Optimal Priority-First Search Decoding:

In this year, we derived two decoding metrics that can be used by the priority first search algorithm [1][2]. Both metrics will lead to the optimal maximum-likelihood decoding. The difference is that the first metric f_1 can be computed on-the-fly, and will therefore cause much less delay in the decoding. For the evaluation of the second metric f_2, however, one needs to know *all* received symbols, but its computational complexity is much less than that of f_1. Continuing the derivation from (1) based on $\mathbb{B}_k^T\mathbb{B}_k = \mathbb{G}_k$ for $1 \le k \le M$, we establish that:

$$\hat{\boldsymbol{b}} = \arg\min_{\boldsymbol{b}\in\mathcal{C}} \frac{1}{2}\sum_{k=1}^{M}\sum_{m=1}^{Q+P}\sum_{n=1}^{Q+P}\left(-w_{m,n,k}b_{(k-1)Q-P+m}b_{(k-1)Q-P+n}\right)$$

where for $1 \le m, n \le Q+P$,

$$w_{m,n,k} = \sum_{i=0}^{P}\sum_{j=0}^{P}\delta_{i,j,k}\text{Re}\{\tilde{y}_{m+i,k}\tilde{y}^*_{n+j,k}\},$$

$$\tilde{\boldsymbol{y}}_k \triangleq [\mathbf{0}_{1\times P}\ \boldsymbol{y}_k^H\ \mathbf{0}_{1\times P}]^H = [\tilde{y}_{1,k}\ \cdots\ \tilde{y}_{Q+2P,k}]^T,$$

and $\delta_{i,j,k}$ is the (*i,j*)-th entry of matrix $\mathbb{D}_k \triangleq \mathbb{G}_k^{-1}$. By adding a constant $\frac{1}{2}\sum_{k=1}^{M}\sum_{m=1}^{Q+P}\sum_{n=1}^{Q+P}|w_{m,n,k}|$ to the decoding criterion, the on-the-fly metric f_1 that suits for the recursive computation of the priority-first search is given by:

$$f_1(b_1,\ldots,b_\ell) = f_1(b_1,\ldots,b_{\ell-1}) + \begin{cases} \alpha_{s,k} - b_\ell\displaystyle\sum_{i=0}^{P}\sum_{j=0}^{P}\delta_{i,j,k}\text{Re}\{\tilde{y}_{s+i,k}\cdot u_{j,k}(b_1,\ldots,b_\ell)\}, \\ \hfill \text{for } P < s \le Q; \\ \alpha_{r,k} - b_\ell\displaystyle\sum_{i=0}^{P}\sum_{j=0}^{P}\delta_{i,j,k}\text{Re}\{\tilde{y}_{r+i,k}\cdot u_{j,k}(b_1,\ldots,b_\ell)\} \\ +\alpha_{s,k+1} \\ -b_\ell\displaystyle\sum_{i=0}^{P}\sum_{j=0}^{P}\delta_{i,j,k+1}\text{Re}\{\tilde{y}_{s+i,k+1}\cdot u_{j,k+1}(b_1,\ldots,b_\ell)\}, \\ \hfill \text{otherwise,} \end{cases}$$

where $s \triangleq [(\ell+P-1) \bmod Q]+1,\ r \triangleq s+Q,\ k \triangleq \max\{\lceil \ell/Q\rceil, 1\}$,

$$\alpha_{s,k} \triangleq \sum_{n=1}^{s-1}|w_{s,n,k}| + |w_{s,s,k}|/2,$$

and

$$u_{j,k}(b_1,\ldots,b_{\ell+1}) = u_{j,k}(b_1,\ldots,b_\ell) + \left(b_\ell\tilde{y}^*_{s+j,k} + b_{\ell+1}\tilde{y}^*_{s+j+1,k}\right)/2$$

with initial values $f_1(b_1,\ldots,b_\ell)=0$ for $\ell=0$, and $u_{j,k}(b_1,\ldots,b_{(k-1)Q-P+1})=0$ for $0\le j\le P$ and $1\le k\le M$. The low-complexity decoding metric f_2 is given by

$$f_2(b_1,\ldots,b_\ell)=f_1(b_1,\ldots,b_\ell)+h(b_1,\ldots,b_\ell),$$

where

$$h(b_1,\ldots,b_\ell)\triangleq\begin{cases}\displaystyle\sum_{m=s+1}^{Q+P}\alpha_{m,k}-\sum_{m=s+1}^{Q+P}|v_{m,k}(b_1,\ldots,b_\ell)|-\beta_{s,k} & \text{for } P<s\le Q;\\ \displaystyle\sum_{m=s+1}^{Q+P}\alpha_{m,k+1}-\sum_{m=s+1}^{Q+P}|v_{m,k+1}(b_1,\ldots,b_\ell)|-\beta_{s,k+1} & \\ \displaystyle\quad+\sum_{m=r+1}^{Q+P}\alpha_{m,k}-\sum_{m=r+1}^{Q+P}|v_{m,k}(b_1,\ldots,b_\ell)|-\beta_{r,k} & \text{otherwise},\end{cases}$$

where s, r and k are defined the same as for $f_1(\cdot)$,

$$v_{m,k}(b_1,\ldots,b_\ell)=v_{m,k}(b_1,\ldots,b_{\ell-1})+w_{s,m,k}b_\ell,$$

and

$$\beta_{s,k}\;=\;\beta_{s-1,k}-\sum_{n=s+1}^{Q+P}|w_{s,n,k}|-\frac{1}{2}|w_{s,s,k}|$$

with initial values $v_{m,k}(b_1,\ldots,b_{(k-1)Q-P+1})=0$ and $\beta_{0,k}=\sum_{m=1}^{Q+P}\alpha_{m,k}$.

2.1.4 Achievement:

The channel parameters $\boldsymbol{h}$ in simulations is zero-mean complex-Gaussian distributed with $E[\boldsymbol{h}\boldsymbol{h}^H]=(1/(P+1))\mathbb{I}_{P+1}$. Note again that $\boldsymbol{h}$ is assumed an unknown constant vector at the system design stage; hence, the system designer does not know whether $\boldsymbol{h}$ is zero-mean complex-Gaussian distributed. Figure 2 then simulates three half-rate codes over frequency selective channels of memory order 1, in which the channel coefficients vary independently in every 15 symbols. The three codes are identified by (28, 14)($Q=29$), (28, 14)($Q=15$) and CS(14, 7), which respectively represent the constructed (28, 14) code with design parameter $Q=29$ (i.e., assuming at the design stage, the channel coefficients remain constant at least during the entire decoding block $L=N+P=28+1=29$), the constructed (28, 14) code with design parameter $Q=15$ (i.e., assuming the channel coefficients vary in every 15 symbols at the design stage), and the computer-searched (hence, structureless) (14, 7) code that minimizes the union bound derived based on the assumption that the channel taps remain constant during the decoding block (i.e., $Q=L=N+P=14+1=15$, which is exactly the simulated channel).

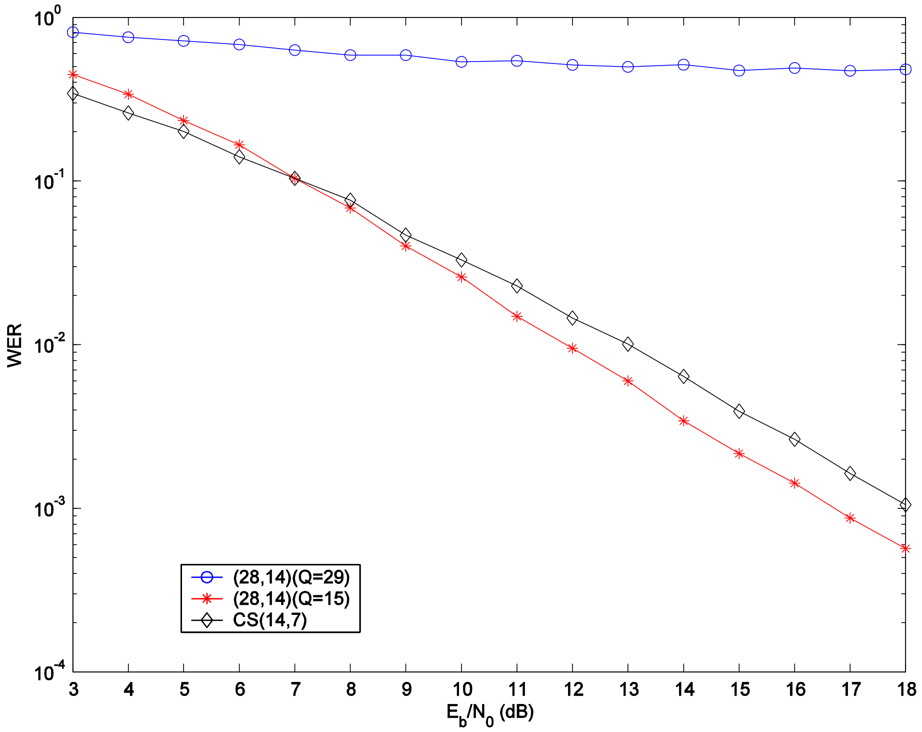

Figure 2: Word error rates (WERs) for the (28, 14)(Q=29) code, the (28, 14)(Q=15) code and the CS(14, 7) code over channels of memory order 1, whose coefficients varying independently in every 15 symbols.

As anticipated, (28, 14) $(Q = 29)$ code seriously degrades in performance since its corresponding assumption at the design stage does not match the characteristic of the true simulated channel. This suggests that the assumption that the channel coefficients remain constant in a coding block is very critical in the code design, and should be made with caution. A striking result from Figure 2 is that the constructed (28, 14)$(Q = 15)$ code performs markedly better than the CS(14, 7) code at medium-to-high signal-to-noise ratios, despite that the CS(14, 7) code is the computer-optimized code specifically for the simulated channel. This suggests that when the channel memory order and varying characteristic are *prior* known (i.e., P and Q), performance gain can be obtained by enhancing the inter-Q-block correlation, and the system favors a longer code design. In Table 1, we summarize the decoding complexity for the (28, 14) $(Q = 15)$ code simulated in Figure 2, measured by the average number of node expansions per information bit. It shows, as previously mentioned, that the decoding metric f_2 requires less decoding efforts than the on-the-fly decoding metric f_1.

The performance of our constructed code can be further (slightly) improved if the codewords are selected uniformly from all feasible code design parameters $(c_1, c_2, \cdots, c_M) \in \{-1, 0, 1\}^M$. For example, select only half (i.e., 2^{13}) of the codewords according to $c_1 = 0$ and $c_2 = -1$ for the (28, 14)$(Q = 15)$ code, and pick the remaining half of the codewords from those binary sequences satisfying

$$\begin{cases} \mathbb{B}_1^T\mathbb{B}_1 = \begin{bmatrix} Q & c_1 \\ c_1 & Q-1 \end{bmatrix} \\ \mathbb{B}_k^T\mathbb{B}_k = \begin{bmatrix} Q & c_k \\ c_k & Q \end{bmatrix} \text{ for } 2 \le k \le M-1 \\ \mathbb{B}_M^T\mathbb{B}_M = \begin{bmatrix} N-(M-1)Q & c_M \\ c_M & N-[(M-1)Q-1]^+ \end{bmatrix} \end{cases}$$

with $c_1 = 0$ and $c_2 = 1$. This however will slightly increase the decoding complexity. The trade-off between selecting codewords from fixed $(c_1, \ldots, c_M)$ or multiple $(c_1, \ldots, c_M)$'s is thus evident.

SNR	3dB	4dB	5dB	6dB	7dB	8dB	9dB	10dB	11dB	12dB	13dB	14dB	15dB
f_1	1658	1367	1074	899	701	593	488	448	356	309	277	244	232
f_2	766	625	482	392	321	254	219	177	149	133	121	104	92
f_1/f_2	2.2	2.2	2.2	2.3	2.2	2.3	2.5	2.4	2.4	2.3	2.3	2.3	2.5

Table 1: Average number of node expansions per information bit for the (28, 14)(Q=15) code simulated in Figure 1.

2.2 Reviews of the work in the second year:

As the number of mobile users as well as the requirement for data rate is rapidly increasing in modern communication systems, the base stations are gradually evolved from macro-cell-based to micro-cell-based. In particular, the service range of a macro-cell base station may be partitioned into several small ones, which are in turn served by several micro-cell base stations[4]. As such, in order to maintain the seamless data transmission, signals from multiple base stations are required to provide softer handover functionality. On the other hand, the demand for mobility is also increased recently, resulting in a more frequent softer handover. Thus, in order to provide high mobility and high data transmission rate simultaneously, we consider in this project a novel system, in which the data is encoded and distributed over N base stations such that the receiver can decode data successfully as long as a certain portion of signals (at least K) from N base stations are received. Since the channel model only requires at least K among N signals are received, it is named the *(N,K)-limited access channel*. In the second year of this project, we analyzed the channel capacity of (N,K)-limited access channel with arbitrary channel models and proposed an fast algorithm to evaluate the optimal power allocation which achieves the channel capacity.

2.2.1 The system model:

As shown in Figure 3, we consider a system that consists of N parallel channels, in which only a certain portion of channel outputs are guaranteed to be successfully received at the receiver end. The receiver however does not *a priori* know which outputs will be nullified or blocked, nor does the receiver have the knowledge of the statistics of these blockages. We can realize this assumption by introducing a set of auxiliary multiplicative constants $s_1, s_2, \ldots, s_N$ to the channel outputs, where the *i*th channel output is nullified when being multiplied by $s_i = 0$, and remains when the multiplicative constant s_i is equal to 1. It is assumed that by monitoring the channel activities, the receiver can perfectly tell the value of $\boldsymbol{s} = [s_1, s_2, \ldots, s_N]^T$. Furthermore, $\boldsymbol{s}$ will remain constant within a codeword transmission period but may vary for different codeword blocks. The receiver will then decode the information based on the receptions $[\boldsymbol{s} \circ \boldsymbol{Y}_1, \boldsymbol{s} \circ \boldsymbol{Y}_2, \ldots, \boldsymbol{s} \circ \boldsymbol{Y}_n]$ if at least K out of N components of vector $\boldsymbol{s}$ are equal to one, where $\boldsymbol{Y}_i \triangleq [Y_{1,i}, Y_{2,i}, \ldots, Y_{N,i}]^T$ are the channel symbols received at time instance *i*, *n* is the codeword length, and operator "$\circ$" denotes the matrix Hadamard product[5]. Conversely, the receiver will give up the decoding if $\sum_{i=1}^{N} s_i < K$. We thus refer to this channel model as an (*N,K*)-limited access channel.

In this setting, we are interested in the optimal power allocation $\boldsymbol{p}^* = [p_1^*, p_2^*, \ldots, p_N^*]^T$ such that the minimum input-output mutual information subject to $\sum_{i=1}^{N} s_i \geq K$ is maximized. This quantity is generally regarded as the achievable rate under which the decoding error can be made arbitrarily small.

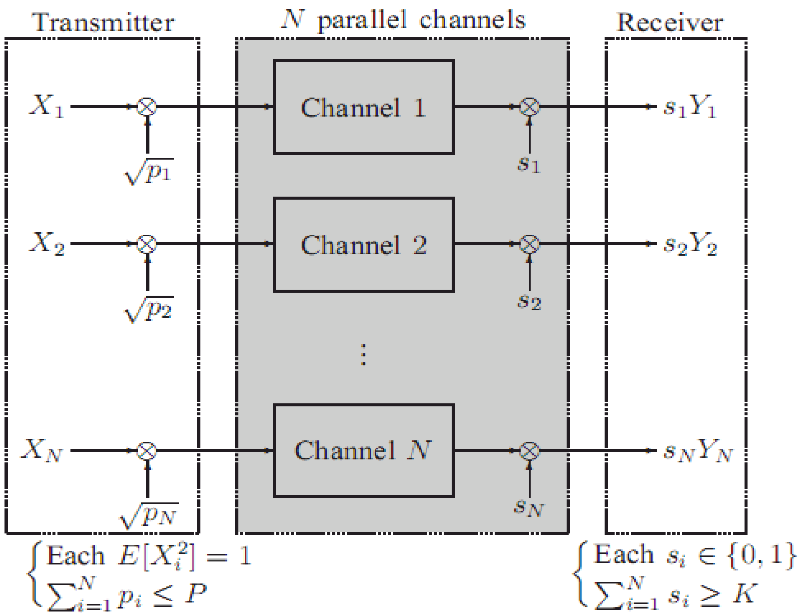

Figure 3: System model for an (N, K)-limited access channel.

Under the system model, the input-output mutual information can be in principle represented by

$$I(\sqrt{\boldsymbol{p}} \circ \boldsymbol{X}; \boldsymbol{s} \circ \boldsymbol{Y})$$

where $I(\cdot;\cdot)$ is the mutual information function and $\sqrt{\boldsymbol{p}} \triangleq [\sqrt{p_1}, \sqrt{p_2}, \ldots, \sqrt{p_N}]^T$. Here, we overload the notation by denoting the channel output vector corresponding to one channel usage by $\boldsymbol{Y} = [Y_1, Y_2, \ldots, Y_N]^T$, and likewisely denote by $\boldsymbol{X} = [X_1, X_2, \ldots, X_N]^T$ the channel input vector for a single channel usage. The achievable rate that guarantees a vanishing decoding error subject to $\sum_{i=1}^N s_i \geq K$ is therefore optimistically

$$\max_{\boldsymbol{X}} \max_{\left\{\boldsymbol{p}\in\Re_+^N : \sum_{i=1}^N p_i \leq P\right.} \min_{\left\{\boldsymbol{s}\in\{0,1\}^N : \sum_{i=1}^N s_i \geq K\right\}} I(\sqrt{\boldsymbol{p}} \circ \boldsymbol{X}; \boldsymbol{s} \circ \boldsymbol{Y}) \tag{2}$$

Where $\Re_+$ is the set of nonnegative real numbers. If the parallel channels are independent in the sense that

$$\Pr(\boldsymbol{Y} | \sqrt{\boldsymbol{p}} \circ \boldsymbol{X}) = \prod_{i=1}^N \Pr(Y_i | \sqrt{p_i} X_i) \tag{3}$$

then the independence bound for mutual information gives that

$$I(\sqrt{\boldsymbol{p}} \circ \boldsymbol{X}; \boldsymbol{s} \circ \boldsymbol{Y}) \leq \sum_{i=1}^N I(\sqrt{p_i} X_i; s_i Y_i) = \sum_{i=1}^N s_i I(\sqrt{p_i} X_i; Y_i)$$

where the last equality follows from that s_i is either 1 or 0. We can therefore focus on the optimal power allocation for independent input distributions, if the channel transition probability satisfies (3).

We next denote for convenience $f_i(p) \triangleq I(\sqrt{p_i} X_i; Y_i)$ for $1 \leq i \leq N$, and make the following assumption on these mutual information functions.

Assumption 1: For $1 \leq i \leq N$, $f_i(p)$ is continuous and strictly increasing for $p \geq 0$, and its first derivative, i.e.,

$$f_i'(p) \triangleq \frac{\partial f_i(p)}{\partial p}$$

exists and is continuous and strictly decreasing in $p \geq 0$, where we define $f_i'(0) \triangleq \lim_{p\downarrow 0} f_i'(p)$.

This assumption will be adopted as a premise in the following analysis. Under *Assumption 1*, it is clear that $f_i(p)$ is a strictly concave function of p with initial value $f_i(0) = I(0; Y_i) = 0$. Together with the fact that $f_i(p) \geq 0$ for $p \in \Re_+$, we can replace the two inequality constraints in (2) by their equality counterparts as

$$\begin{aligned} &\max_{\left\{\boldsymbol{p}\in\Re_+^N : \sum_{i=1}^N p_i \leq P\right.} \min_{\left\{\boldsymbol{s}\in\{0,1\}^N : \sum_{i=1}^N s_i \geq K\right\}} \textstyle\sum_{i=1}^N s_i \cdot f_i(p_i) \\ &= \max_{\left\{\boldsymbol{p}\in\Re_+^N : \sum_{i=1}^N p_i = P\right.} \min_{\left\{\boldsymbol{s}\in\{0,1\}^N : \sum_{i=1}^N s_i = K\right\}} \textstyle\sum_{i=1}^N s_i \cdot f_i(p_i) \end{aligned} \tag{4}$$

for a given $\boldsymbol{X}$ that validates *Assumption 1*.

In the next section, we will show that under *Assumption 1*, the maximization-minimization problem in (3) becomes algorithmically tractable (cf. *Theorem 2*).

2.2.2 Analysis of The optimal power allocation

In this section, the analysis for the optimization problem in (4) is presented.

For $K = 1$, (4) can be simplified to

$$\max_{\left\{\boldsymbol{p}\in\Re_+^N:\sum_{i=1}^N p_i=P\right\}} \min\{f_1(p_1), f_2(p_2), \ldots, f_N(P_N)\}.$$

It is thus straightforward that the optimal power allocation $\boldsymbol{p}^*$ satisfies

$$f_1(p_1^*) = f_2(p_2^*) = \cdots = f_N(p_N^*)$$

For $K = N$, the maximization-minimization power allocation problem reduces to one that only requires a maximization computation because $s_1 = s_2 = \ldots = s_N = 1$. Therefore, one can apply the Lagrange multipliers technique and Karush-Kuhn-Tucker (KKT) condition to find the optimal power allocation [6]. However, for $1 < K < N$, there does not exist a straight technique for this maximization-minimization problem. Nevertheless, we can find a necessary condition for the optimal power allocation such that the labor of examining all possible $\binom{N}{K}$ combinations of $\boldsymbol{s}$ satisfying $\sum_{i=1}^N s_i \geq K$ can be reduced as indicated in the next lemma.

Lemma 1: The optimal power allocation $\boldsymbol{p}^*$ for an (*N,K*)-limited access channel satisfies

$$f_{a_1}(p_{a_1}^*) \leq f_{a_2}(p_{a_2}^*) \leq \cdots \leq f_{a_K}(p_{a_K}^*) = f_{a_{K+1}}(p_{a_{K+1}}^*) = \cdots = f_{a_N}(p_{a_N}^*)$$

for some permutation $a_1, a_2, \ldots, a_N$ of sequence $1, 2, \ldots, N$.

An immediate implication of *Lemma 1* is that we can distinguish the optimal power allocation for an (*N,K*)-limited access channel into *K* disjoint cases. In other words, the condition

$$\max_{1\leq i\leq \ell-1} f_{a_i}(p_{a_i}^*) < f_{a_\ell}(p_{a_\ell}^*) = f_{a_{\ell+1}}(p_{a_{\ell+1}}^*) = \cdots = f_{a_N}(p_{a_N}^*) \tag{5}$$

is valid exactly for one value of ℓ in $\{1, 2, \ldots, K\}$. As a result, if the index set

$$\mathbb{A} \triangleq \{a_\ell, a_{\ell+1}, \cdots, a_N\}$$

in which their respective mutual information function values are equal to $\max_{1\leq i\leq N} f_i(p_i^*)$ is identified in advance, the maximization-minimization power allocation problem is simplified to a maximization problem as

$$\max_{\boldsymbol{p}\in\mathcal{P}(\mathbb{A})} \left\{ \sum_{i\notin\mathbb{A}} f_i(p_i) + (K - N + |\mathbb{A}|) \max_{1\leq j\leq N} f_j(p_j) \right\} \tag{6}$$

where

$$\mathcal{P}(\mathbb{A}) \triangleq \left\{ \boldsymbol{p} \in \Re_{+}^{N} : \begin{array}{l} (i) \sum_{i=1}^{N} p_i = P \\ (ii) f_i(p_i) < \max_{1\le j\le N} f_j(p_j) \text{ for } i \notin \mathbb{A} \\ (iii) f_i(p_i) = \max_{1\le j\le N} f_j(p_j) \text{ for } i \in \mathbb{A} \end{array} \right\}. \tag{7}$$

However, the direct identification of $\mathbb{A}$ without knowing $\boldsymbol{p}^*$ in advance is in general a challenged. The opposite, i.e., identifying $\mathbb{A}$ after determining $\boldsymbol{p}^*$, is more straightforward. In order to resolve the optimization problem, we propose in the following subsections to first determine the best power allocation $\boldsymbol{p}^\diamond$ corresponding to a *conjectured* maximal-mutual-information index set, denoted by $\mathbb{B}$. Then, we will examine afterwards whether this conjecture is the optimal one or not based on some condition we will establish later. In case the conjectured $\mathbb{B}$ only achieves a suboptimal power allocation, a new round of maximization computation and follow-up examination will be launched based on a newly generated $\mathbb{B}$. Since the established condition will help identifying one index that is not in $\mathbb{A}$ at each round, the process can hopefully stop after $N - |\mathbb{A}| + 1$ iterations after which $\boldsymbol{p}^*$ is obtained.

A. Determination of the best power allocation $\boldsymbol{p}^\diamond$ corresponding to a given index set $\mathbb{B}$

Based on a given index set $\mathbb{B}$, we transform the maximization-minimization problem into

$$\sup_{\boldsymbol{p}\in\mathcal{P}(\mathbb{B})} \left\{ \sum_{i\notin\mathbb{B}} f_i(p_i) + (K - N + |\mathbb{B}|) \max_{1\le j\le N} f_j(p_j) \right\} \tag{8}$$

where $\mathcal{P}(\mathbb{B})$ is defined the same as (7) except that $\mathbb{A}$ is replaced with $\mathbb{B}$. Since the given $\mathbb{B}$ may not be the optimal index set $\mathbb{A}$, the solution $\boldsymbol{p}^\diamond$ of the optimization problem defined in (8) could be at the boundary of $\mathcal{P}(\mathbb{B})$ in the sense that

$$f_i(p_i^\diamond) = \max_{1\le j\le N} f_j(p_j^\diamond) \quad \text{for some } i \notin \mathbb{B}.$$

For this reason, we use *supremum* instead of *maximum* in (8).

We next show that this inequality constraint can be relaxed by means of the incorporation of the *aggregate mutual information function* that transforms the *N*-dimensional power allocation problem into an equivalent $N - |\mathbb{B}| + 1$-dimensional one.

Definition 1: The *aggregate mutual information* function $F_{\mathbb{B}}$ with respect to a sequence of mutual information functions $\{f_i\}_{i\in\mathbb{B}}$ is defined through its inverse function as

$$F_{\mathbb{B}}^{(\texttt{inv})}(y) \triangleq \sum_{i\in\mathbb{B}} f_i^{(\texttt{inv})}(y) \quad \text{for } y \ge 0 \tag{9}$$

provided that all the inverse functions exist (which is guaranteed by *Assumption 1*).

A graphical illustration of the aggregate mutual information function for $\mathbb{B}=\{1,2,3\}$ is given in Figure 4. In this figure, it is clear that

$$F_{\mathbb{B}}^{(\texttt{inv})}=f_1^{(\texttt{inv})}(y)+f_2^{(\texttt{inv})}(y)+f_3^{(\texttt{inv})}(y)=p_1+p_2+p_3$$

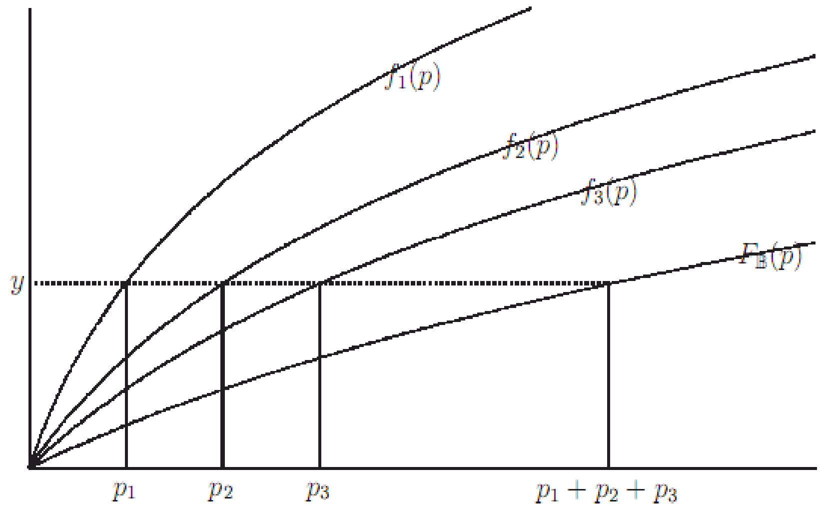

Figure 4: Graphical illustration of the aggregate mutual information function when $f_i(p)=\log(1+p/\sigma_i^2)$ and $\sigma_i^2=i$ for $i\in\mathbb{B}=\{1,2,3\}$.

As a specific example, if $f_i(p)=\log(1+p/\sigma_i^2)$ for some $\sigma_i^2>0$ and $1\le i\le 3$, then

$$F_{\mathbb{B}}(p)=\log\left(1+\frac{p}{\sigma_1^2+\sigma_2^2+\sigma_3^2}\right).$$

In terms of the aggregate mutual information function, we can simplify the constraints in $\mathcal{P}(\mathbb{B})$ in the following lemma.

Lemma 2: Fix an index set $\mathbb{B}$. The solution $\boldsymbol{p}^\diamond$ of the optimization problem in (8) satisfies

$$p_i^\diamond=\begin{cases} q_i^\diamond & \text{for } i\notin\mathbb{B} \\ f_i^{(\texttt{inv})}(F_{\mathbb{B}}(q_{\mathbb{B}}^\diamond)) & \text{for } i\in\mathbb{B}\end{cases} \tag{10}$$

where the $N-|\mathbb{B}|+1$-dimensional vector $\boldsymbol{q}^\diamond$ is the solution of the optimization problem below:

$$\sup_{\boldsymbol{q}\in\mathcal{Q}(\mathbb{B})}\{\textstyle\sum_{i\notin\mathbb{B}} f_i(q_i)+(K-N+|\mathbb{B}|)F_{\mathbb{B}}(q_{\mathbb{B}})\} \tag{11}$$

where

$$\mathcal{Q}(\mathbb{B})\triangleq\left\{\boldsymbol{q}=\ (\text{list of } q_i\forall i\notin\mathbb{B}, q_{\mathbb{B}})\ \in\Re_+^{N-|\mathbb{B}|+1}:\ \begin{array}{l}(i)\sum_{i\notin\mathbb{B}} q_i+q_{\mathbb{B}}=P\\ (ii)f_i(q_i)<F_{\mathbb{B}}(q_{\mathbb{B}}) \text{ for } i\notin\mathbb{B}\end{array}\right\}.$$

In addition, $\boldsymbol{q}^\diamond\in\mathcal{Q}(\mathbb{B})$ if, and only if, $\boldsymbol{p}^\diamond\in\mathcal{P}(\mathbb{B})$.

By the reduction of constraints down to two in $\mathcal{Q}(\mathbb{B})$ in *Lemma 2*, we can further proceed to show that the inequality constraint in $\mathcal{Q}(\mathbb{B})$ is redundant in case $\boldsymbol{q}^{\diamond} \in \mathcal{Q}(\mathbb{B})$ as summarized in *Theorem 1*.

Theorem 1: Given that $\boldsymbol{q}^{\diamond} \in \mathcal{Q}(\mathbb{B})$, the maximize $\boldsymbol{q}^{\diamond}$ for (11) is equal to the maximize $\tilde{\boldsymbol{q}}$ of the problem below:

$$\max_{\boldsymbol{q} \in \tilde{\mathcal{Q}}(\mathbb{B})} \{\textstyle\sum_{i \notin \mathbb{B}} f_i(q_i) + (K - N + |\mathbb{B}|) F_{\mathbb{B}}(q_{\mathbb{B}})\} \tag{12}$$

where

$$\tilde{\mathcal{Q}}(\mathbb{B}) \triangleq \{\boldsymbol{q} \in \Re_{+}^{N-|\mathbb{B}|+1} : \textstyle\sum_{i \notin \mathbb{B}} q_i + q_{\mathbb{B}} = P\}$$

We conclude this subsection by pointing out that the maximization computation in (12) is now performed over the usual single power-sum constraint, and hence can be solved by the Lagrange multipliers technique and KKT condition by treating $(K - N + |\mathbb{B}|) F_{\mathbb{B}}(q_{\mathbb{B}})$ as the mutual information function of an auxiliary aggregate channel. Based on the result in *Theorem 1*, we are ready to present the algorithmic approach that helps identifying the optimal maximal-mutual-information index set $\mathbb{A}$ and the optimal power allocation $\boldsymbol{p}^{*}$.

B. Determination of the Optimal Maximal-Mutual-Information Index Set $\mathbb{A}$ and the Optimal Power Allocation $\boldsymbol{p}^{*}$

For an (*N*,*K*)-limited access channel, there are possibly $\sum_{\ell=1}^{K} \binom{N}{\ell-1}$ candidate index sets for the choices of $\mathbb{B}$ in *Theorem 1*, and it may be time-consuming to perform the optimization computation for (12) for each of them. The next theorem then shows that this time-consuming maximization labor can be reduced to only $N - |\mathbb{A}| + 1$.

Theorem 2: The optimal maximal-mutual-information index set $\mathbb{A}$ as well as the optimal power allocation $\boldsymbol{p}^{*}$ can be obtained through the following algorithmic procedure:

Step 1. Initialize $M = 1$ and $\mathbb{B}_1 = \{1, 2, \ldots, N\}$.

Step 2. Obtain the maximize $\tilde{\boldsymbol{q}}_M$ for (12) by setting $\mathbb{B} = \mathbb{B}_M$, and calculate

$$\tilde{\boldsymbol{p}}_M = [\tilde{p}_{M,1}, \tilde{p}_{M,2}, \cdots, \tilde{p}_{M,N}]^T$$

corresponding to the obtained $\tilde{\boldsymbol{q}}_M$ and the given $\mathbb{B}_M$ through an assignment similar to (10).

Step 3. Assign $\mathbb{B}_{M+1} = \mathbb{B}_M \setminus \{j_M\}$ where j_M is an index in $\mathbb{B}_M$ that satisfies

$$f'_{j_M}(\tilde{p}_{M,j_M}) = \min_{i \in \mathbb{B}_M} f'_i(\tilde{p}_{M,i}) \tag{13}$$

(If there are more than one indices satisfying (13), just pick up any one of them as j_M.)

Step 4. If

$$(K-M)F'_{\mathbb{B}_{M+1}}(\sum_{i\in\mathbb{B}_{M+1}}\tilde{p}_{M,i}) \le f'_{j_M}(\tilde{p}_{M,j_M}) \tag{14}$$

then set $\mathbb{A}=\mathbb{B}_M$ and $\boldsymbol{p}^*=\tilde{\boldsymbol{p}}_M$ and stop the algorithm; otherwise, set $M=M+1$ and go to Step 2.

We would like to point out that the algorithm in *Theorem 2* will stop when (usually before) M reaches K because (14) trivially holds when $M=K$. This coincides with the definition of $\mathbb{A}$ in (5) that at most $K-1$ indices are outside $\mathbb{A}$.

Theorem 2 indicates that given the first derivative of the marginal mutual information function $f_i(p)=I(\sqrt{p}X_i;Y_i)$ being positive, strictly decreasing and continuous in p for every $1\le i\le N$ (i.e., *Assumption 1*), we can determine the optimal power allocation $\boldsymbol{p}^*$ for a spatially independent (*N,K*)-limited access channel with input $\sqrt{\boldsymbol{p}}\circ\boldsymbol{X}$ by performing $N-|\mathbb{A}|+1$ maximizations in the sense of (12).

2.2.3 Achievement:

For (N,K)-limited access channels with arbitrary inputs, the capacity formula is derived as a maximization-minimization problem. We then analyze the maximization-minimization problem to get two properties as shown in *Lemma 1* and *Lemma 2*. According to these two properties and the definition of aggregate mutual information, we then simplify the maximization-minimization problem to a simple maximization problem with only one single power-sum constraint. Based on the simple maximization problem with single power-sum constraint, we propose an algorithm to find the optimal power allocation $\boldsymbol{p}^*$ by $N-|\mathbb{A}|+1$ time-consuming maximization labor.

3. 報告內容(第三年)

3.1 Introduction:

In the second year of this project, we have proposed an algorithm of finding the optimal power allocation for general channels with limited access constraint. Following the proposed algorithm, in this year we further establish that when channel disturbances, in addition to independence, are reduced to being additive with distributions scaled from a common random variable, the optimal power allocation can be directly obtained from a *two-phase water-filling* process if the arbitrary inputs are given by the respective component variables in an independent and identical distributed (i.i.d.) random vector, multiplying by the square root of the allocated power. The two-phase water-filling interpretation then hints that the degree of "noisiness" for a general (possibly, non-additive and non-Gaussian) limited access channel might be identified by composing the derivative of the mutual information function with its inverse.

3.2 The system model:

Although Gaussians are generally appropriate noise models for physical additive channels, experimental measurement indicates that the noises in certain environments are by no means Gaussian distributed [7][8][9]. As such, in the third year of this project, we consider additive noise of the same family in (N, K)-limited access channels.

By additive noises of the same family, we mean that the relationship between channel inputs and outputs can be characterized by

$$Y_i = \sqrt{p_i} X_i + \sigma_i Z_i \quad \text{for } 1 \le i \le N \tag{15}$$

where $\{X_i\}_{i=1}^N$ and $\{Z_i\}_{i=1}^N$ are both i.i.d. complex random variables with unit second moments, and they are independent from each other; the system model is shown as Figure 5.

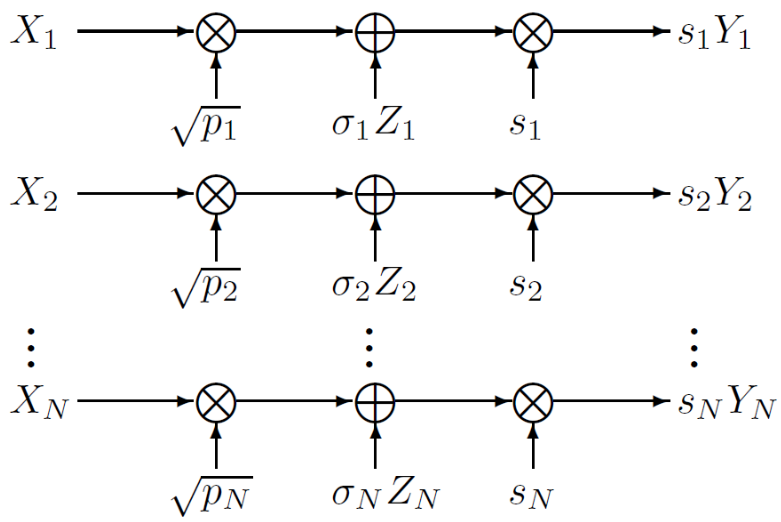

Figure 5: System model for an (N, K)-limited access channel with additive noise of the same family, where $E[|X_i|^2] = E[|Z_i|^2] = 1$, $s_i \in \{0,1\}$ for $1 \le i \le N$, $\sum_{i=1}^N s_i \ge K$ and $\sum_{i=1}^N p_i \le P$.

We then restrict our attention only to the case that Z_i is a continuous random variable because *Assumption 1*(at page 8) may fail when both X_i and Z_i are discrete. Notably, X_i often takes values in a finite alphabet (e.g., $\{\pm 1\}$) in practice. Specifically, when the intersection of two sets $\left\{\sqrt{p_i}x + \sigma_i z : P_{Z_i}(z) > 0\right\}$ and $\left\{\sqrt{p_i}\tilde{x} + \sigma_i z : P_{Z_i}(z) > 0\right\}$ is empty for every $x \ne \tilde{x}$ with $P_{X_i}(x) > 0$ and $P_{X_i}(\tilde{x}) > 0$, we have

$$f_i(p_i) = I(\sqrt{p_i} X_i; Y_i) = H(\sqrt{p_i} X_i) = H(X_i)$$

where $H(X_i)$ is the entropy of the channel input X_i [10]. This implies that in a discrete system, $f_i(p_i)$ can be equal to its maximum value $H(X_i)$ almost everywhere in p_i, in which case *Assumption 1* is unquestionably violated.

Observe that for continuous additive noises,

$$
\begin{aligned}
I(\sqrt{p_i}X_i;Y_i) &= h(Y_i) - h(Y_i|\sqrt{p_i}X_i) \\
&= h(Y_i) - h(\sqrt{p_i}X_i + \sigma_i Z_i|\sqrt{p_i}X_i) \\
&= h(\sigma_i\tilde{Y}_i) - h(\sigma_i Z_i) \\
&= h(\tilde{Y}_i) - h(Z_i) \\
&= I\left(\frac{\sqrt{p_i}}{\sigma_i}X_i;\tilde{Y}_i\right)
\end{aligned} \tag{16}
$$

where $h(\cdot)$ is the differential entropy function [10], and (16) follows from the independence between X_i andZ_i, and $\tilde{Y}_i \triangleq (\sqrt{p_i}/\sigma_i)X_i + Z_i$. This immediately yields

$$
f_i(p_i) = g\left(\frac{p_i}{\sigma_i^2}\right) \text{ for every } 1 \le i \le N \tag{17}
$$

with

$$
g(\rho) \triangleq I(\sqrt{\rho}X_i;\sqrt{\rho}X_i + Z_i). \tag{18}
$$

Assumption 1 thus reduces to the single condition that function g is continuous and strictly increasing, and its first derivative exists and is continuous and strictly decreasing.

3.3 The optimal power allocation for additive noise of the same family:

Based on this system setting, we show in the next theorem that the optimal power allocation $\boldsymbol{p}^*$ follows a *two-phase water-filling scheme*} Specifically, in the first phase (which we refer to as the *noise-power re-distribution phase*), the least $N-K$ noise powers among $\{\sigma_i^2\}_{i=1}^N$ will be first poured as *noise water* into a tank consisting of K interconnected vessels with solid base heights equal to the remaining K noise powers and with widths of unit length as shown in Figure 6(b). Afterwards those W vessels either with water inside or with solid base height equal to the water surface level will be subdivided into $N-K+W$ vessels of rectangular shape with the same heights (as the water surface level) and with widths in proportion to their noise powers (but the total volume remaining unchanged). As such, a tank with N vessels of proper heights and widths (corresponding to N channels) is ready for the second phase as exemplified in Figure 6(c). It is worth mentioning that after the first phase, the optimal maximal-mutual-information index set $\mathbb{A}$ has already been identified and consists of the channel indices corresponding to the aforementioned W vessels and the least $N-K$ noise powers (hence, $|A| = W + N - K$).

In the second phase (which we refer to as the *signal-power allocation phase*), the heights of vessel bases will be first either *lifted* or possibly *lowered* according to total signal power P and function g as well as their current heights as shown in Figure 6(e). What follows, as exemplified in Figure 6(f), is the usual water-filling power allocation scheme. The pre-adjustment of base heights before water filling can be viewed as preparation for these vessels to be “capable” of supporting the water that is going to be poured in with amount P. As a result, the volume of water ended up in each vessel is exactly the power that should be allocated. Notably, for the special case that the noises $\{Z_i\}_{i=1}^N$ are complex Gaussian distributed, the heights of vessel bases can never be lowered in the pre-adjustment step; hence, a *mercury-filling* scheme before water pouring has been proposed to materialize the lifting of heights of vessel bases [11]. However, since the adjustment of heights of vessel bases generally can be in both *up* and *down* directions, the use of the name *mercury/water filling* may induce that the vessel bases should be lifted under general non-Gaussian additive noises; hence, we simply use the conventional name of water-filling in this work.

Theorem 3: Suppose that the information transmitted over an (N,K)-limited access channel is corrupted by additive noises of the same family characterized by (15) and the mutual information function $g(\rho)$ defined in (18) satisfies *Assumption 1*. Assume without loss of generality that

$$\sigma_1^2 \geq \sigma_2^2 \geq \cdots \geq \sigma_N^2.$$

Then, the optimal maximal-mutual-information index set $\mathbb{A}$ is given by

$$\mathbb{A} = \{\ell, \ell+1, \cdots, N\} \tag{19}$$

where

$$\ell \triangleq \min\left\{ i \in \{1,2,\cdots,K\} \middle| \sigma_i^2 \leq \tilde{\sigma}_K^2 \text{ for every } 1 \leq i \leq K \right\} \tag{20}$$

and $\tilde{\sigma}_i^2 \triangleq \sigma_i^2 + [\lambda - \sigma_i^2]^+$ for $1 \leq i \leq K$ with λ chosen to satisfy $\sum_{i=1}^K [\lambda - \sigma_i^2]^+ = \sum_{i=K+1}^N \sigma_i^2$, and $[y]^+ \triangleq \max\{0, y\}$. The optimal power allocation $\boldsymbol{p}^*$ can therefore be obtained from $\boldsymbol{q}^*$ through an assignment similar to (10), where $\boldsymbol{q}^*$ is the maximizer for (12) with $\mathbb{B}$ equal to the above $\mathbb{A}$. In other words,

$$p_i^* = \begin{cases} q_i^* & \text{for } 1 \leq i < \ell \\ \frac{\sigma_i^2}{\sum_{j=\ell}^N \sigma_j^2} \cdot q_{\mathbb{A}}^* & \text{for } \ell \leq i \leq N \end{cases} \tag{21}$$

with

$$q_i^* = \left\{ \begin{array}{ll} \sigma_i^2 \cdot g'^{(\texttt{inv})}\left(\nu\,\sigma_i^2\right) & \text{if } g'(\infty) < \nu\sigma_i^2 < g'(0) \\ 0 & \text{if } \nu\sigma_i^2 \geq g'(0) \end{array} \right\} \text{ for } 1 \leq i < \ell \tag{22}$$

and

$$q_{\mathbb{A}}^* = \left(\sum_{j=\ell}^N \sigma_j^2 \right) \cdot g'^{(\texttt{inv})} \left(\nu \frac{\sum_{j=\ell}^N \sigma_j^2}{K - \ell + 1} \right) \tag{23}$$

where $g'^{(\texttt{inv})}$ is the inverse function of the first derivative g' of function g, and ν is chosen such that

$$\sum_{i=1}^{\ell-1} q_i^* + q_{\mathbb{A}}^* = P. \quad (24)$$

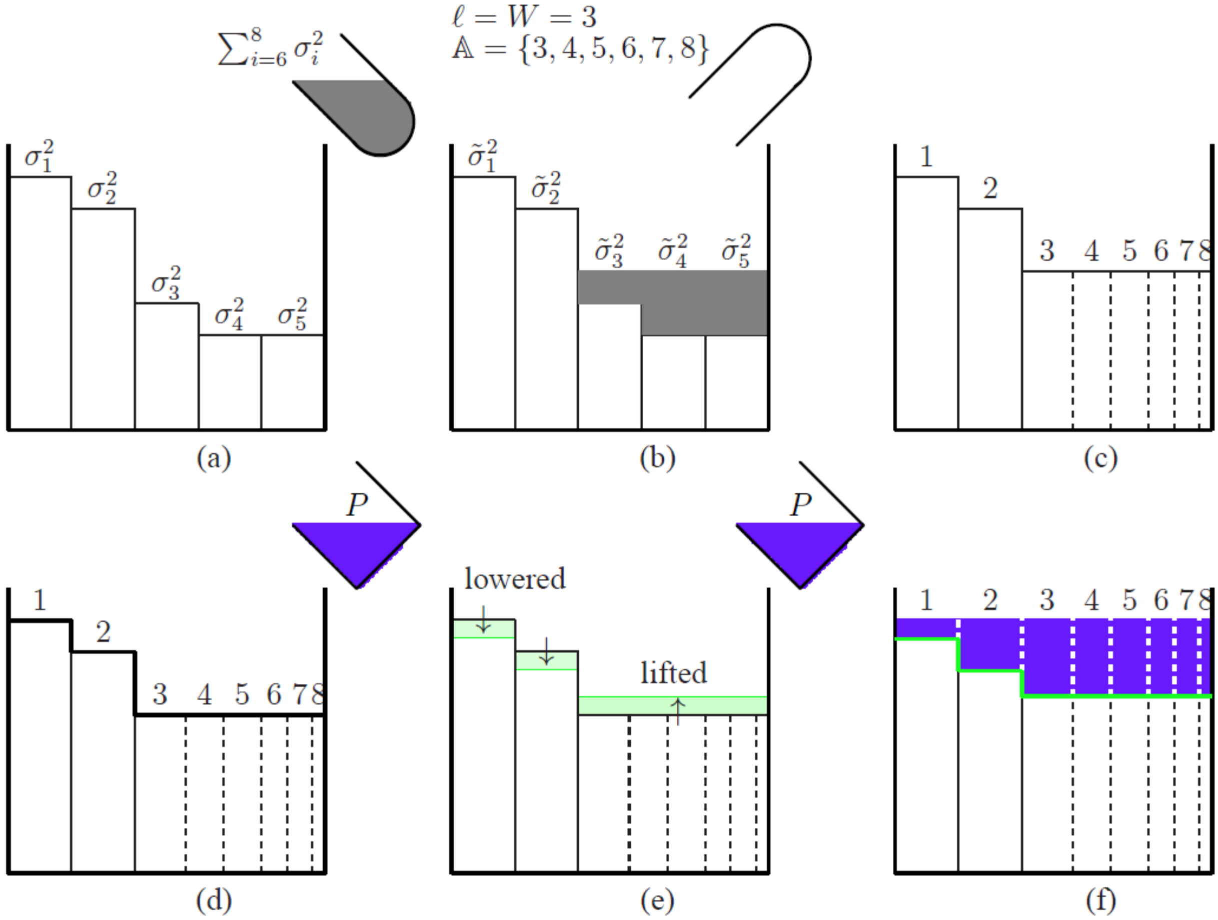

Figure 6: The graphical interpretation of the optimal *two-phase water-filling* power allocation for an $(8,5)$ -limited access channel with independent additive noises characterized by (15). In this figure, $[\sigma_1^2, \sigma_2^2, \cdots, \sigma_8^2] = [8,7,4,3,3,2,2,1]$. Subfigures (a), (b) and (c) correspond to the *noise-power redistribution* phase, while subfigures (d), (e) and (f) illustrate the *signal-power allocation* phase.

Several remarks can be made based on *Theorem 3*.

- First, it can be extended from *Theorem 3* that as long as $\mathbb{A}$ is pre-determined, the maximization labor can always be reduced down to one. In the special case that the noises are additive and originated from the same family (as considered in this section), we can directly determine $\mathbb{A}$ in terms of (20).
- Secondly, when $\ell = 1$ (equivalently, $\mathbb{A} = \{1, 2, \ldots, N\}$), $\boldsymbol{p}^*$ can be determined without any maximization labor since we immediately have $q_{\mathbb{A}}^* = P$ by (24). In such a case, the optimal power allocation follows the *equal signal-to-noise ratio* (SNR) principle as

$$\frac{p_i^*}{\sigma_i^2} = \frac{P}{\sum_{j=1}^{N} \sigma_j^2} \quad \text{for every } 1 \le i \le N.$$

- Finally, the validity of *Theorem 3* does not need to be restricted to channels with additive noises of the same family but can be extended to any (N, K)-limited access channel with marginal mutual information functions satisfying (17) for some function g that obeys

Assumption 1. A straightforward example is the flat fading channels with known channel states at the receiver end, characterized by

$$Y_i = (\beta_i H_i)(\sqrt{p_i} X_i) + \sigma_i Z_i \quad \text{for } 1 \le i \le N \tag{25}$$

where $\{H_i\}_{i=1}^N$ is i.i.d. with unit second moment, and is independent of the channel input and additive noise. We then obtain $f_i(p_i) = g(\beta_i^2 p_i / \sigma_i^2)$ with $g(\rho) = I(\sqrt{\rho} X_i; \sqrt{\rho} H_i X_i + Z_i | H_i)$. *Theorem 3* thus can be used to establish the optimal power allocation by treating σ_i^2 / β_i^2 as the new noise power level.

An exemplified illustration of the two-phase water-filling scheme is depicted in Figure 6. Details are given below.

<*The noise-power re-distribution phase*>

Fig. 6(a) Set K vessels with widths of unit length and with base height of the ith vessel being σ_i^2 for $1 \le i \le K$. (Note that we assume $\sigma_1^2 \ge \sigma_2^2 \ge \cdots \ge \sigma_N^2$.)

Fig. 6(b) Pour in the "*noise water*" of amount $\sum_{j=K+1}^N \sigma_j^2$ and set $\tilde{\sigma}_i^2$ as the new water level of vessel i for $1 \le i \le K$. Let ℓ be the smallest integer among $\{1, 2, \ldots, K\}$ such that $\sigma_i^2 \le \tilde{\sigma}_K^2$ (cf. (20)). Assign $\mathbb{A} = \{\ell, \ell+1, \ldots, N\}$ and $W = K - \ell + 1$.

Fig. 6(c) Sub-divide the space of the last W vessels (i.e., $K-W+1, K-W+2, \ldots, K$) into $W + (N-K)$ new vessels of rectangular shape with base height the same as the water surface level and widths in proportion to σ_i^2 for $\ell \le i \le N$.

< *The signal-power allocation phase* >

Fig. 6(d) Retain the N vessels from the previous phase.

Fig. 6(e) Adjust the base height of the ith vessel to

$$L_i(\nu) \triangleq \begin{cases} \sigma_i^2 \cdot G(\nu \sigma_i^2) & \text{for } 1 \le i < \ell \\ \tilde{\sigma}_K^2 \cdot G(\nu \tilde{\sigma}_K^2) & \text{for } \ell \le i \le N \end{cases} \tag{26}$$

where ν is the parameter chosen in *Theorem 3* according to (24), and

$$G(\zeta) \triangleq \begin{cases} \frac{1}{\zeta} - g'^{(\mathtt{inv})}(\zeta) & \text{if } g'(\infty) < \zeta < g'(0) \\ \frac{1}{g'(0)} & \text{if } \zeta \ge g'(0). \end{cases}$$

Fig. 6(f) Pour in the "*signal water*" of amount P. Then the volume of water in the ith vessel is the optimal power p_i^* to be allocated for channel i.

3.4 Implications from the optimal power allocation:

Theorem 2 indicates that the sequence of candidate maximal-mutual-information index sets $\mathbb{B}_1, \mathbb{B}_2, \mathbb{B}_3, \ldots$ can be identified via the determination of $j_1, j_2, j_3, \ldots$. In a sense, this sequence can be regarded as sorting the channels in their descending degrees of "noisiness," which can be supported by the result from *Theorem 3*, where the sequence of $j_1, j_2, j_3, \ldots$ coincides with $\sigma_{j_1}^2 \geq \sigma_{j_2}^2 \geq \sigma_{j_3}^2 \geq \cdots$.

For a general (N, K)-limited access channel in which the noises are not necessarily additive or scaled from the same family, can one identify such sequence through their mutual information functions? The next theorem may provide a guide along this direction of thinking.

Theorem 4: For a general (N, K)-limited access channel, if

$$f'_{k_1}\left(f_{k_1}^{(\mathtt{inv})}(y)\right) \leq f'_{k_2}\left(f_{k_2}^{(\mathtt{inv})}(y)\right) \leq \cdots \leq f'_{k_N}\left(f_{k_N}^{(\mathtt{inv})}(y)\right) \quad \text{for all } y \geq 0$$

then $j_M = k_M$ for $M = 1, 2, 3, \ldots$.

Here, regardless of the original goal of the determination of optimal power allocation, *Theorem 4* (as an extension from *Theorem 3*) proposes a way to compare the degree of "noisiness" of general channels via their mutual information functions. For the additive noise channels of the same family, we have

$$f'_i\left(f_i^{(\mathtt{inv})}(y)\right) = \frac{1}{\sigma_i^2} g'\left(g^{(\mathtt{inv})}(y)\right).$$

Hence, the proposed ordering coincides with the general impression that the larger the σ_i^2, the noisier the ith channel is considered to be. To simplify the notation, we drop the parentheses between f'_i and $f_i^{(\mathtt{inv})}$ in the sequel.

For channels other than additive noise of the same family, there could be no apparent winner between any two channels in the sense of $\{f'_i f_i^{(\mathtt{inv})}\}_{i=1}^N$. In other words, it could happen that

$$f'_i f_i^{(\mathtt{inv})}(y_1) > f'_j f_j^{(\mathtt{inv})}(y_1) \quad \text{but} \quad f'_i f_i^{(\mathtt{inv})}(y_2) < f'_j f_j^{(\mathtt{inv})}(y_2)$$

for two distinct y_1 and y_2 and two distinct i and j. As such, the sequence of $j_1, j_2, j_3, \ldots$ will become a function of the total signal power P. However, if a certain condition is satisfied, the pre-identification of the degrees of channel noisiness is still possible at two extreme situations: $P \to 0$ and $P \to \infty$, which we will respectively refer to as the low- and high-power regimes in later discussion.

Lemma 3:

1. If

$$\limsup_{y \downarrow 0} \left(f'_i f_i^{(\mathtt{inv})}(y) - f'_j f_j^{(\mathtt{inv})}(y)\right) \leq 0 \quad \text{for every } 1 \leq i < j \leq N \tag{27}$$

   then $j_i = i$ in the low-power regime, where sign function $\operatorname{sgn}(\rho)$ is equal to either 1, 0 or -1 depending on whether $\rho > 0$, $\rho = 0$ or $\rho < 0$.

2. If

$$\limsup_{y \uparrow \min\{\omega_i, \omega_j\}} \operatorname{sgn}\left(f_i' f_i^{(\mathtt{inv})}(y) - f_j' f_j^{(\mathtt{inv})}(y)\right) \le 0 \quad \text{for every } 1 \le i < j \le N \tag{28}$$

then $j_i = i$ in the high-power regime, provided that $\lim_{p\to\infty} f_i'(p) = 0$ for $1 \le i \le N$, where $\omega_i \triangleq \lim_{p\to\infty} f_i(p)$.

Since the input alphabet is usually finite for channels of practical interest, we have $\omega_i \triangleq \lim_{p\to\infty} f_i(p) \le H(X_i) < \infty$. This immediately validates the premise, i.e., $\lim_{p\to\infty} f_i'(p) = 0$, for condition (28) implying $j_i = i$ in the high-power regime. In other words, $\lim_{p\to\infty} f_i'(p) = 0$ is true for all finite-input channels. There however exists a certain kind of channels where $\omega_i = \infty$ while $\lim_{p\to\infty} f_i'(p) = 0$. An example is the Gaussian-input AWGN channel for which $f_i(p) = \log(1 + p/\sigma_i^2)$. We would like to emphasize that the inference regarding (28) still remains valid for channels with unbounded mutual information as long as $\lim_{p\to\infty} f_i'(p) = 0$.

Conditions (27) and (28) in *Lemma 3* involve the examination of the limit supremum of function differences. The following corollary shows that their validity can be guaranteed by comparing the limiting behaviors of individual functions.

Corollary 1:

1. The validity of (27) for an (i, j) pair is certain if one of the three conditions below is satisfied:

$$\begin{cases} f_i'(0) < f_j'(0) \\ f_i'(0) = f_j'(0) \text{ and } f_i^{(}0) < f_j^{(}0) \\ (\exists\, \delta > 0)\ f_i'(p) \le f_j'(p) \text{ for } 0 < p < \delta \end{cases} \tag{29}$$

2. The validity of (28) for an (i, j) pair is certain if

$$\omega_i = \lim_{p\to\infty} f_i(p) < \omega_j = \lim_{p\to\infty} f_j(p). \tag{30}$$

According to the above discussions, we can identify the degree of noisiness for general channel easily by the sufficient conditions provided in *Theorem 4*, *Lemma 3* and *Corollary 1*.

4. Reference

[1] Y. S. Han and P.-N. Chen, "Sequential decoding of convolutional codes," *The Wiley Encyclopedia of Telecommunications*, edited J. Proakis, John Wiley and Sons, Inc., 2002.

[2] Y. S. Han, C. R. P. Hartmann and C.-C. Chen, "Efficient priority-first search maximum-likelihood soft-decision decoding of linear block codes," *IEEE Trans. Inform. Theory*, vol. 39, no. 5, pp. 1514–1523, Sep. 1993.

[3] P.-N. Chen *et al.*, *Final Report of NSC 93 −2213 −E −009 −050*, May, 2009.

[4] H. Holma and A. Toskala, *WCDMA for UMTS: HSPA Evolution and LTE*, 4th ed.. Chichester: UK, Wiley, 2007.

[5] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.

[6] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[7] K. L. Blackard, T. S. Rappaport, and C. W. Bostain, "Measurements and models of radio frequency impulsive noise for indoor wireless communications," *IEEE J. Sel. Areas Commun.,* vol. 11, no. 7, pp. 991-1001, Sep. 1993.

[8] M. G. Sanchez, A. V. Alejos, and I. Cuinas, "Urban wide-band measurements of the UMTS electromagnetic environment," *IEEE Trans. Veh. Technol.,* vol. 53, no. 4, pp. 1014-1022, Jul. 2004.

[9] M. Zimermann and K. Dostert, "Analysis and modeling of impulsive noise in broad-band powerline communications," *IEEE Trans. Electromagn. Compat.,* vol. 44, no. 1, pp. 249-258, Feb. 2002.

[10] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* New York: Wiley, 1991.

[11] A. Lozano, A. M. Tulino, and S. Verdu, "Optimal power allocation for parallel Gaussian channels with arbitrary input distributions," *IEEE Trans. Inf. Theory,* vol. 52, no. 7, pp. 3033-3051, Jul. 2006.

4. 計畫成果自評

In this three-years project, we have investigated several scenarios of codes designing for non-coherent detection system that combines channel estimation and error correction. This design can directly construct a code of any desired code length and code rate, of which the performance is shown to be comparable to the best computer-searched code for the channels simulated. For the designing and analysis of the novel (N, K)-limited access system, we have derived the channel capacity and proposed a fast algorithm of finding optimal power allocation to achieve the capacity. Following the proposed algorithm, the optimal power allocation can be obtained by a two-phase water-filling process when the channel model is additive noise of the same family. From the interpretation of two-phase water-filling, we further define the degree of noisiness for general channels. The works for the novel (N, K)-limited access system will appear in *IEEE Transactions on Information Theory* and was presented in part at the *2011 International Symposium on Information Theory*.