

Introduction to OFDM and its PAPR Drawback

Dr. Prapun Suksompong

หลักการเบื้องต้นของ OFDM

OFDM ซึ่งย่อมาจาก Orthogonal Frequency Division Multiplexing เป็นระบบที่มีการวิจัยอย่างแพร่หลายในปัจจุบัน มีการนำไปประยุกต์ใช้กับการสื่อสารไร้สายหลายประเภท ข้อดีเบื้องต้นของระบบ OFDM คือ เป็นระบบที่สามารถส่งผ่านข้อมูลด้วยอัตราความเร็วสูงเมื่อเทียบกับระบบอื่นๆ อีกทั้งยังใช้แบนด์วิดท์ที่มีอยู่อย่างมีประสิทธิภาพซึ่งจะได้กล่าวถึงในส่วนต่อไป ในส่วนนี้เราจะอธิบายหลักการทำงานเบื้องต้นของ OFDM โดยเริ่มจากระบบที่ไม่ซับซ้อนเพื่อให้เข้าใจได้ง่าย หลังจากนั้นเราจะชี้ให้เห็นถึงปัญหาและการแก้ปัญหาโดยการเพิ่มเติมคุณลักษณะของ OFDM ลงไปในระบบ

การส่งข้อมูลโดยใช้คลื่นพาห่เดี่ยว (Single Carrier Transmission)

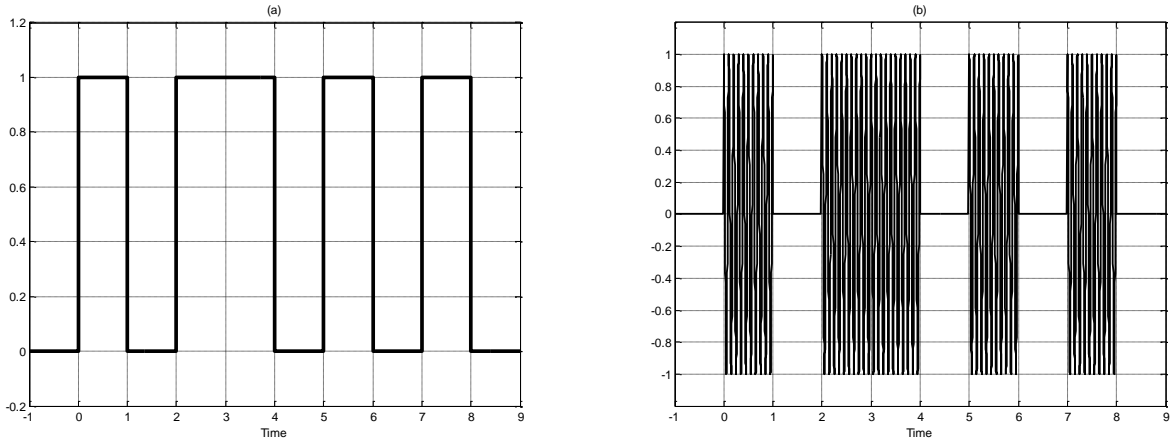
เทคนิคการส่งสัญญาณแบบ OFDM เป็นรูปแบบของการส่งข้อมูลแบบขนานโดยอาศัยหลายคลื่นพาห่ (Multi-Carrier Transmission) ซึ่งเราจะนำเสนอโดยเปรียบเทียบกับวิธีการส่งข้อมูลโดยใช้คลื่นพาห่เดี่ยว (Single Carrier Transmission) ในการส่งข้อมูลโดยใช้คลื่นพาห่เดี่ยวนั้น การส่งชุดข้อมูล (block of symbols) ที่ประกอบด้วย s_1, s_2, \dots, s_N สามารถทำได้โดยการส่ง pulse $p(t)$ ที่เปลี่ยนแปลงไปตาม s_1, s_2, \dots, s_N เรียงกันไปตามลำดับ โดยสามารถเขียนเป็นสมการได้ดังนี้

$$s(t) = \sum_{k=0}^{N-1} s_k p(t - kT)$$

โดย T เป็นเวลาที่ใช้ในการส่งแต่ละสัญลักษณ์ข้อมูล (symbol interval) ตัวอย่างของ pulse $p(t)$ ที่ง่ายที่สุดก็คือ pulse รูปสี่เหลี่ยม (rectangular pulse) ซึ่งเขียนสมการได้เป็น

$$p(t) = 1_{[0,T)}(t) = \begin{cases} 1, & t \in [0, T) \\ 0, & \text{otherwise.} \end{cases}$$

เราจะเห็นว่าข้อมูล s_1 นั้นเมื่อนำไปคูณกับ pulse $p(t)$ แล้วจะอยู่ระหว่างเวลา $t \in [0, T)$ ในแกนเวลา ส่วนข้อมูล s_2 นั้นจะอยู่ระหว่างเวลา $t \in [T, 2T)$ ดังนั้น เมื่อใช้ pulse รูปสี่เหลี่ยมข้างต้นสัญญาณที่ได้จะไม่มีการซ้อนทับกันในแกนเวลา รูปที่ 1a แสดงตัวอย่างของสัญญาณ $s(t)$ สำหรับข้อมูล $(s_1, s_2, \dots, s_8) = (1, 0, 1, 1, 0, 1, 0, 1)$



รูปที่ 1: (a) ตัวอย่างของสัญญาณ $s(t)$ เมื่อข้อมูลเป็น $(s_1, s_2, \dots, s_8) = (1, 0, 1, 1, 0, 1, 0, 1)$ (b) ตัวอย่างของสัญญาณ $x(t)$ ที่เกิดจากการ modulate สัญญาณ $s(t)$

ก่อนที่สัญญาณ $s(t)$ ดังกล่าวจะถูกส่งออกไปจากเครื่องส่ง สัญญาณจะถูกแปลงความถี่ด้วยสัญญาณพาห้โดยกระบวนการที่เรียกว่า modulation ถ้าเราวิเคราะห์สัญญาณในแกนความถี่ (frequency domain) ผลของการ modulation ก็คือสัญญาณจะย้ายแถบความถี่ของตัวเองจากรอบ ๆ ความถี่ศูนย์ไปยังรอบ ๆ ความถี่ของสัญญาณพาห้ ดังนั้นเราจึงเรียกสัญญาณที่ถูก modulate แล้วว่า **สัญญาณแบนด์พาส (bandpass signal)** สำหรับสัญญาณที่ยังไม่ได้ผ่านการ modulation เราจะเรียกว่า **สัญญาณเบสแบนด์ (baseband signal)**

เราสามารถเขียนสมการของสัญญาณแบนด์พาสได้ดังต่อไปนี้

$$x(t) = \text{Re}\{s(t)e^{j2\pi f_c t}\}$$

โดยที่ f_c เป็นความถี่ของคลื่นพาห้ ถ้าเราให้ $S(f)$ เป็น spectrum ของสัญญาณเบสแบนด์ $s(t)$ ผลของการ modulation ในแกนความถี่ก็คือ spectrum $X(f)$ ของสัญญาณ bandpass $x(t)$ สามารถเขียนได้เป็น

$$X(f) = \frac{1}{2}(S(f - f_c) + S^*(-f - f_c))$$

จะเห็นได้ว่า spectrum $S(f)$ ของ $s(t)$ ได้ย้ายไปอยู่ที่ $\pm f_c$ นอกจากนี้ถ้าสัญญาณเบสแบนด์ $s(t)$ เป็นจำนวนจริงเสมอ สมการของสัญญาณแบนด์พาสข้างต้นจะเหลือเพียง

$$x(t) = s(t)\cos(2\pi f_c t)$$

ส่วน spectrum $X(f)$ ก็จะเป็น

$$X(f) = \frac{1}{2}(S(f - f_c) + S(f + f_c))$$

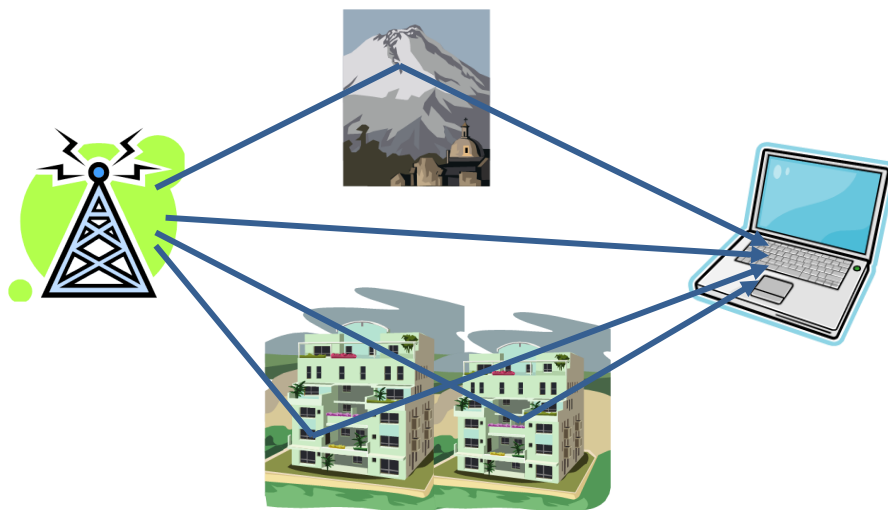
รูปที่ 1b แสดงตัวอย่างของสัญญาณ bandpass ที่ได้จากสัญญาณ baseband ในรูปที่ 1a

สัญญาณที่ถูก modulate แล้วจะถูกส่งผ่านไปยังช่องสัญญาณ (Channel) ในส่วนถัดไปเราจะอธิบายลักษณะของช่องสัญญาณสำหรับการสื่อสารแบบไร้สาย

การสื่อสารไร้สายและการจางหายแบบหลายเส้นทาง (Wireless communication and Multipath fading)

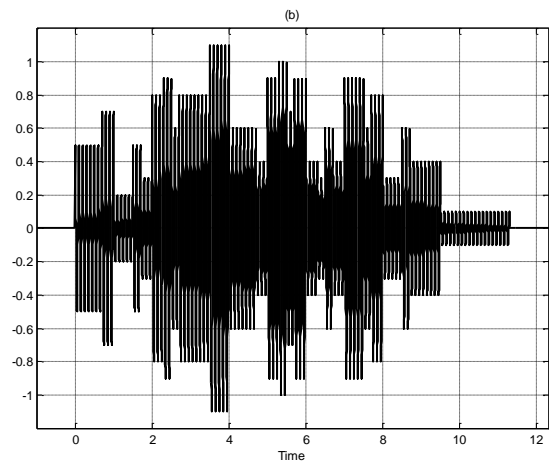
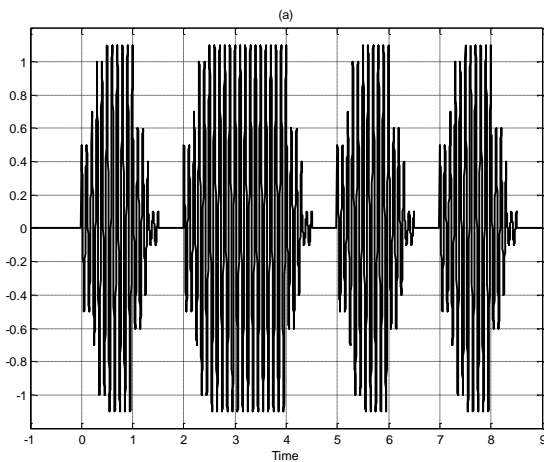
การสื่อสารแบบไร้สายได้กลายเป็นส่วนสำคัญในการดำรงชีวิตประจำวัน ในขณะเดียวกัน ความต้องการส่งข้อมูลเป็นปริมาณมากในเวลารวดเร็วได้เพิ่มสูงขึ้นเรื่อย ๆ อย่างไรก็ตาม การสื่อสารผ่านช่องสัญญาณแบบไร้สายนั้นมีข้อจำกัดมากมาย คุณลักษณะของช่องสัญญาณสามารถเปลี่ยนแปลงตามองค์ประกอบต่าง ๆ ได้ เช่น สิ่งแวดล้อมทางกายภาพ ความเร็วในการเคลื่อนที่ของอุปกรณ์ภาครับหรือส่ง และ อุณหภูมิ เป็นต้น นอกจากนี้ยังมีสัญญาณรบกวน (noise) ซึ่งเราควบคุมได้ยาก ในส่วนนี้เราจะกล่าวถึง การจางหายแบบหลายเส้นทางของช่องสัญญาณ

ในการสื่อสารแบบไร้สาย สัญญาณที่ถูกส่งออกมาจะกระจายออกเป็นหลายทิศทาง ดังนั้นจึงมีการสูญเสียของสัญญาณสูง เมื่อคลื่นสัญญาณเดินทางผ่านตัวกลางก็จะมีการลดทอน (Attenuation) ของสัญญาณเกิดขึ้น นอกจากนี้สัญญาณที่กระจายออกนอกเส้นทางยังสามารถสะท้อนกลับมายังเครื่องรับได้อีกดังรูปที่ 2 สัญญาณที่สะท้อนกลับมาเหล่านี้เดินทางมาถึงเครื่องรับล่าช้ากว่าสัญญาณที่มาจากเส้นทางตรงเพราะระยะทางที่เพิ่มขึ้น เราจะเรียกเวลาการเดินทางที่เพิ่มขึ้นนี้ว่า **ค่าหน่วงเวลา (excess delay)** โดยปกติแล้วสัญญาณสะท้อนจะมีมากกว่าหนึ่งสัญญาณ โดย ขนาด และ ค่าหน่วงเวลาของแต่ละสัญญาณ ก็จะแตกต่างกันไป เราเรียกการเดินทางของสัญญาณแบบหลายเส้นทางนี้ว่า **multipath propagation** และเรียกสัญญาณเหล่านี้ว่า **multipath waves**



รูปที่ 2: Multipath propagation

เมื่อเกิด multipath propagation สัญญาณที่เครื่องรับได้รับก็จะมีอาการบิดเบี้ยว (smearing) ไปจากสัญญาณที่ถูกส่งออกมาจากเครื่องส่งดังแสดงในรูปที่ 3 ปัญหานี้เราเรียกว่า การจางหายแบบหลายเส้นทาง หรือ multipath fading สังเกตว่าพลังงานบางส่วน of สัญญาณแรกซึ่งควรจะถูกจำกัดอยู่ระหว่างเวลา 0 ถึง T ได้ล้าเข้าไปอยู่ในส่วนของสัญญาณที่สอง ปัญหานี้เรียกว่า การสอดแทรกระหว่างสัญลักษณ์ (Inter-symbol Interference: ISI)



รูปที่ 3: (a) Multipath fading เมื่อ T (symbol interval) มีขนาดใหญ่เมื่อเทียบกับค่าหน่วยเวลา จะเห็นได้ว่าสัญญาณไม่ได้เปลี่ยนไปมากนักจากสัญญาณ $x(t)$ ในรูปที่ 1 (b) Multipath fading เมื่อ T (symbol interval) มีขนาดเล็กเคียงกับค่าหน่วยเวลา จะเห็นได้ว่าสัญญาณนั้นแตกต่างไปจาก สัญญาณ $x(t)$ ในรูปที่ 1 อย่างมาก

เราสามารถจำลองลักษณะของช่องสัญญาณที่มี multipath fading ได้ด้วยสมการทางคณิตศาสตร์ที่อธิบายผลตอบสนองอิมพัลส์ของช่องสัญญาณ (channel impulse response) ดังต่อไปนี้

$$h(t) = \sum_{i=0}^v \beta_i \delta(t - \tau_i)$$

เมื่อสัญญาณ $x(t)$ เดินทางผ่านช่องสัญญาณ สัญญาณ $r(t)$ ที่ปรากฏที่เครื่องรับจะเป็น

$$x(t) * h(t) + n(t) = \sum_{i=0}^v \beta_i x(t - \tau_i) + n(t)$$

โดยที่ $n(t)$ คือสัญญาณรบกวน และ เครื่องหมาย * แสดงถึงการประสาน (convolution) ของสัญญาณ $x(t)$ กับ $h(t)$ เมื่อเราวิเคราะห์สมการดังกล่าว เราจะเห็นว่ามีการรวมของสัญญาณ $\beta_i x(t - \tau_i)$ ซึ่งก็คือ สัญญาณ $x(t)$ ที่มีค่าหน่วยเวลา τ_i และขนาดถูกลดทอนไปเป็น β_i นั่นเอง ใน รูปที่ 3a เราใช้

$$h(t) = 0.5\delta(t) + 0.2\delta(t - 0.2T) + 0.3\delta(t - 0.3T) + 0.1\delta(t - 0.5T)$$

ส่วนในรูปที่ 3b เราใช้

$$h(t) = 0.5\delta(t) + 0.2\delta(t - 0.7T) + 0.3\delta(t - 1.5T) + 0.1\delta(t - 2.3T)$$

ข้อสังเกตอีกอย่างจากรูปที่ 3 ก็คือปัญหา multipath fading จะมีน้อยมากถ้า symbol interval T มีขนาดใหญ่เมื่อเทียบกับค่าหน่วยเวลา อย่างไรก็ตาม ความต้องการส่งข้อมูลอย่างรวดเร็วทำให้ไม่สามารถเพิ่มขนาดของ T ได้โดยตรงเพราะจะทำให้อัตราการส่งข้อมูลลดลง การแก้ปัญหา multipath fading นี้สามารถทำได้โดยการปรับแต่งช่องสัญญาณ (Equalization) ซึ่งยุ่งยากและซับซ้อน อีกวิธีหนึ่งที่สามารถหลีกเลี่ยงปัญหา multipath fading และเป็นหลักการพื้นฐานที่ถูกนำมาใช้ใน OFDM ก็คือการส่งข้อมูลโดยแบ่งข้อมูลไปยังหลายแถบความถี่ซึ่งจะได้กล่าวถึงในส่วนถัดไป

Multi-carrier transmission

ดังที่ได้อธิบายในส่วนที่แล้ว ปัญหา multipath fading นั้นจะมีผลกระทบน้อยมากถ้าเราส่งสัญญาณด้วยค่า T ที่มีขนาดใหญ่ แต่ความต้องการส่งข้อมูลอย่างรวดเร็วทำให้ไม่สามารถลดค่า T ได้โดยตรงเพราะจะทำให้อัตราการส่งข้อมูลลดลงตามไปด้วย การส่งข้อมูลโดยใช้หลายคลื่นพาห้เป็นเทคนิคหนึ่งซึ่งหลีกเลี่ยงปัญหา multipath fading ได้ โดยใช้หลักการพื้นฐานที่ไม่ได้แตกต่างไปจากหลักการในการส่งข้อมูลโดยใช้คลื่นพาห้เดียวที่กล่าวมาข้างต้นเลย หากแต่มีการแบ่งข้อมูลออกเป็นหลาย ๆ ส่วนขนานกันไป (parallel data stream) โดยแต่ละส่วนย่อย ถูกส่งออกไปด้วยคลื่นพาห้ย่อย (subcarrier, SC) ที่มีความถี่ต่างกัน เมื่อข้อมูลถูกแบ่งเป็นหลายส่วน แต่ละส่วนจึงไม่จำเป็นต้องมีอัตราการส่งที่สูง ดังนั้นจึงสามารถใช้ T ที่มีขนาดใหญ่ได้ ยิ่งเพิ่มจำนวนของสัญญาณพาห้ย่อยมากขึ้นเท่าใดก็ยิ่งส่งข้อมูลได้มากขึ้นโดยไม่กระทบต่อ T วิธีนี้เรียกว่า การมัลติเพล็กซ์โดยการแบ่งความถี่

(Frequency Division Multiplexing: FDM)

การส่งข้อมูลแบบ FDM นั้นมีข้อจำกัดหลายประการ ประการแรก เนื่องจากการใช้คลื่นพาห้หลายความถี่ เครื่องรับสัญญาณจะต้องทำการแยกสัญญาณที่ถูกส่งมากับแต่ละคลื่นพาห้โดยการกรองสัญญาณ ดังนั้นเราจะต้องจัดให้ความถี่ของคลื่นพาห้ที่อยู่ห่างกันมากพอสมควรเพื่อลดการทับซ้อนของสัญญาณในแกนความถี่ทำให้ไม่สามารถใช้ย่านความถี่ที่มีอยู่อย่างเต็มที่ อีกปัญหาของ FDM เป็นเรื่องของความซับซ้อนของระบบ นั่นคือ ยิ่งเราแบ่งข้อมูลออกมามากส่วนเท่าใด ความซับซ้อนของเครื่องส่งและรับสัญญาณก็ยิ่งมากขึ้นตามไปด้วยเพราะจะต้องรองรับการทำงานภายใต้จำนวนความถี่ของคลื่นพาห้ที่มากขึ้น เครื่องส่งสัญญาณจะต้องมีวงจรกำเนิดความถี่ (oscillator) ที่หลากหลาย ปัญหาทั้งสองประการเป็นอุปสรรคที่สำคัญของระบบแบบ FDM ในส่วนถัดไป เราจะกล่าวถึงการปรับปรุงประสิทธิภาพของ FDM ด้วยการส่งสัญญาณที่ตั้งฉากซึ่งกันและกัน

การตั้งฉาก (Orthogonality)

ในส่วนนี้จะกล่าวถึงการส่งสัญญาณที่ตั้งฉาก (orthogonal) ซึ่งกันและกัน สังเกตว่าอักษร O ในคำย่อ OFDM นั้นมาจากคำว่า orthogonality ซึ่งก็คือ การตั้งฉากที่เราจะกล่าวถึงในส่วนนี้เอง

สมมติว่าเราต้องการส่งข้อมูล s_1, s_2, \dots, s_N เราสามารถทำได้โดยการส่ง pulse $p(t)$ ในแกนเวลาที่เปลี่ยนขนาดไปตาม s_1, s_2, \dots, s_N เรียงกันไปตามลำดับ

$$s(t) = \sum_{k=0}^{N-1} s_k p(t - kT)$$

โดยแต่ละ pulse ใช้เวลา T หน่วยดังที่เราได้อธิบายไปแล้วในส่วนของการส่งข้อมูลโดยใช้คลื่นพาห้เดียว สำหรับการส่งข้อมูลแบบ FDM นั้นในแต่ละช่องเวลาที่มีความถี่ T เราจะส่งข้อมูล s_1, s_2, \dots, s_N ไปพร้อม ๆ กันโดยใช้หลายแถบความถี่ ซึ่งสามารถสรุปเป็นสมการในแกนความถี่ได้ว่า

$$S(f) = \sum_{k=0}^{N-1} s_k p(f - k\Delta f)$$

โดยที่ Δf เป็นระยะห่างระหว่าง spectrum ของสัญญาณย่อย สำหรับในการส่งสัญญาณที่ตั้งฉากกันนั้น สัญญาณ $p(f - k\Delta f)$ จะต้องตั้งฉากซึ่งกันและกันด้วย นั่นคือ สำหรับจำนวนเต็ม k และ j ที่แตกต่างกัน เราต้องการให้

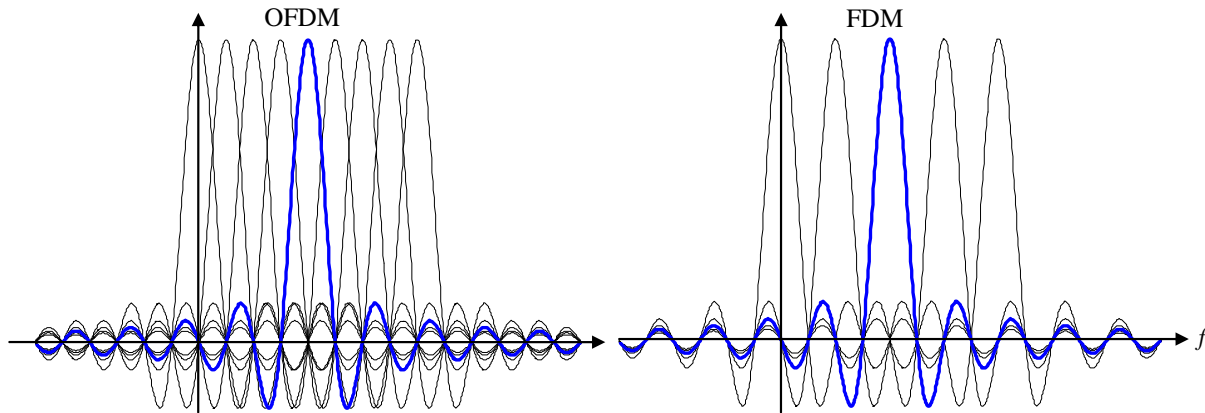
$$\int p(f - k\Delta f) p^*(f - j\Delta f) df = 0$$

เห็นได้ชัดว่า ถ้า $p(f - k\Delta f)$ ไม่ซ้อนทับกับ $p(f - j\Delta f)$ ในแกนความถี่ ผลของการคูณกันจะได้ศูนย์เสมอ ส่งผลให้ค่าที่ได้จากการ integrate เป็นศูนย์ไปด้วย นั่นคือสัญญาณทั้งสองก็จะตั้งฉากกันตามนิยามข้างต้นและไม่รบกวนกัน นี่เป็นหลักการพื้นฐานของ FDM อย่างไรก็ตาม การที่จะแยกส่งสัญญาณให้อยู่ในแถบความถี่ที่ไม่ซ้อนทับกันโดยสิ้นเชิงนั้นทำได้ยากในทางปฏิบัติ และเป็นสาเหตุให้ระบบแบบ FDM ต้องวาง spectrum ของสัญญาณย่อยบนแกนความถี่ให้ห่างกันมาก อีกทั้งยังมีการเพิ่มแถบป้องกัน (guard band) ระหว่าง spectrum ของสัญญาณย่อยอีก ทั้งหมดนี้นำมาซึ่งการสิ้นเปลืองย่านความถี่

ที่จริงแล้วการที่สัญญาณจะตั้งฉากกันนั้น ไม่จำเป็นที่จะต้องหลีกเลี่ยงการทับซ้อนกันในแกนความถี่ ตัวอย่างของชุดสัญญาณที่ตั้งฉากกันแบบนี้ก็คือสัญญาณในรูปของ ฟังก์ชัน sinc ซึ่งใช้

$$p(f) = \frac{1}{\Delta f} \text{sinc}\left(\pi \frac{f}{\Delta f}\right)$$

ดังแสดงในรูปที่ 4 จะเห็นว่า spectrum ของสัญญาณย่อยนั้นซ้อนทับกัน แต่จุดสูงสุด (peak value) ของแต่ละ spectrum ของสัญญาณย่อยหนึ่งจะเป็นจุดที่ spectrum ของสัญญาณย่อยอื่น ๆ มีค่าเป็นศูนย์ทำให้ไม่มีการซ้อนทับกันที่จุดสูงสุดเหล่านี้ สรุปแล้ว OFDM ก็เป็น FDM แบบหนึ่งนั่นเองหากแต่มีการจัดวางคลื่นพาห่อย่อยให้อยู่ใกล้กันที่สุดโดยยังคงความตั้งฉากกันอยู่



รูปที่ 4: Spectrum ของ OFDM และ spectrum ของ FDM

ถ้าเราวิเคราะห์สัญญาณในแกนเวลาโดยใช้ inverse Fourier transform เราจะเห็นว่าสัญญาณที่เป็นฟังก์ชัน sinc ในแกนความถี่นั้นเทียบเท่ากับสัญญาณรูปสี่เหลี่ยมในแกนเวลา การที่ peak ของสัญญาณ sinc นั้นถูกวางไว้ที่ความถี่ $k\Delta f$ นั้นเทียบเท่ากับการคูณด้วย $e^{j2\pi k\Delta f t}$ ดังแสดงในสมการต่อไปนี้

$$p(f - k\Delta f) \xrightarrow{\mathcal{F}^{-1}} \text{rect}_{\left[-\frac{T}{2}, \frac{T}{2}\right]}(t) e^{+j2\pi k\Delta f t}$$

ดังนั้นสัญญาณข้างต้นจะถูกจำกัดอยู่ระหว่างเวลา $t \in \left[-\frac{T}{2}, \frac{T}{2}\right]$ โดยที่ $T = \frac{1}{\Delta f}$ เมื่อนำเอาสัญญาณย่อยมารวมกันเราจะได้

$$s(t) = \sum_{k=0}^{N-1} s_k e^{+j2\pi k\Delta f t}$$

สำหรับเวลา $t \in \left[-\frac{T}{2}, \frac{T}{2}\right]$ แต่ถ้าเราใช้ pulse $p(f)$ อื่นที่ไม่ใช่ ฟังก์ชัน sinc สัญญาณรวมก็จะเป็น

$$s(t) = \sum_{k=0}^{N-1} s_k P(t) e^{+j2\pi k \Delta f t}$$

โดย $P(t)$ คือ inverse Fourier transform ของ $p(f)$

ข้อสังเกต: สัญญาณ $s(t)$ ไม่จำเป็นต้องเริ่มจากเวลา $-T/2$ หากเราต้องการให้สัญญาณถูกจำกัดอยู่ในช่วงเวลา $[t_0, t_0 + T]$ ก็สามารทำได้โดยใช้สัญญาณ

$$s\left(t - \left(t_0 - \frac{T}{2}\right)\right)$$

แทน สัญญาณ $s(t)$

ในส่วนของการรับสัญญาณนั้น ถ้าเราใช้ฟังก์ชัน sinc เป็น pulse $p(f)$ ดังที่ยกตัวอย่างข้างต้น และไม่เกิดการรบกวนภายในช่องสัญญาณ นั่นคือ ถ้าสัญญาณ $r(t)$ ที่ได้รับนั้นเป็นสัญญาณ $s(t)$ โดยตรง เมื่อเราต้องการดึงเอาข้อมูล s_1, s_2, \dots, s_N ออกมาจาก $r(t)$ ก็สามารทำได้โดยการแปลง $r(t)$ ให้อยู่ในแกนความถี่ $R(f)$ ด้วย Fourier transform แล้วชักสัญญาณ (sample) ที่ความถี่

$$f = m\Delta f \text{ โดย } m = 0, 1, \dots, N-1$$

เราจะได้

$$R(m\Delta f) = \sum_{k=0}^{N-1} s_k p((m-k)\Delta f) = \sum_{k=0}^{N-1} s_k \frac{1}{\Delta f} \text{sinc}(\pi(m-k)) = \frac{s_m}{\Delta f}$$

ซึ่งเทียบเท่ากับการรู้ค่าของ s_m เพราะเรารู้ค่าของ Δf อยู่ก่อนแล้ว จะเห็นได้ว่า การที่เราสามารถดึงเอาค่า s_1, s_2, \dots, s_N ออกมาได้อย่างง่ายดายก็เพราะจุด peak ของแต่ละฟังก์ชัน sinc นั้น อยู่ตรงจุดที่ฟังก์ชัน sinc อื่นเป็น 0 ทั้งสิ้น

การนำ IDFT และ DFT มาใช้ใน OFDM

การใช้ orthogonality ในส่วนที่แล้วนั้นทำให้ OFDM สามารถใช้ย่านความถี่ได้อย่างมีประสิทธิภาพเพิ่มขึ้นเกือบเท่าตัว เมื่อเทียบกับ FDM แบบธรรมดา แต่ปัญหาใหญ่อีกประการของ FDM ก็คือการที่เครื่องส่งต้องผลิตสัญญาณพาห้ที่มีความถี่ได้หลากหลาย ซึ่งหาก N มีค่ามากก็เป็นเรื่องที่ทำได้ยาก ในส่วนนี้เราจะอธิบายถึงที่มาของการประมวลสัญญาณด้วย IDFT (inverse discrete Fourier transform) และ DFT (discrete Fourier transform) ใน OFDM ซึ่งทำให้สามารถส่งสัญญาณทั้ง N ความถี่ได้ในเวลาเดียวกัน

ก่อนอื่น เราจะลอง sample สัญญาณ $s(t)$ ที่เวลา

$$t = nT = \frac{n}{\Delta f} \text{ โดย } n = 0, 1, \dots, N-1$$

ซึ่งผลที่ได้ก็คือ

$$s[n] = s(nT) = \sum_{k=0}^{N-1} s_k e^{+j2\pi kn}$$

จะเห็นได้ว่า

$$(s[n])_{n=0}^{N-1} = \text{IDFT} \left\{ (s_k)_{k=0}^{N-1} \right\}$$

ดังนั้น หากไม่เกิดการผิดเพี้ยนของสัญญาณในช่องสัญญาณ เราจะได้

$$r[n] = r(nT) = s[n]$$

และสามารถดึงเอาข้อมูลกลับมาได้โดยใช้ DFT สรุปว่าในการส่งข้อมูลแบบ OFDM นั้น เราจะแปลงข้อมูลจากแบบอนุกรม ซึ่งก็คือข้อมูลที่เรียงกันมาตามลำดับเป็นข้อมูลแบบขนาน แล้วทำการประมวลผลข้อมูลทั้งหมดในเวลาเดียวกันด้วย IDFT จากนั้นจึงส่งออกไปในช่องสัญญาณ ที่เครื่องรับก็จะประมวลผลสัญญาณโดยใช้ DFT เพื่อให้ได้ข้อมูลกลับมามีดังสรุปด้วยแผนผังต่อไปนี้

$$(s_k)_{k=0}^{N-1} \rightarrow \boxed{\text{IDFT}} \rightarrow (s[n])_{n=0}^{N-1} \cdots \rightarrow (r[n])_{n=0}^{N-1} \rightarrow \boxed{\text{DFT}} \rightarrow (R_k)_{k=0}^{N-1} = (S_k)_{k=0}^{N-1}$$

นอกจากนี้ เพราะข้อมูลผ่านการ IDFT ก่อนที่จะถูกส่งออกไป เราจึงเรียกข้อมูลเหล่านี้ว่าสัญลักษณ์ในแกนความถี่ (frequency-domain symbol)

ข้อสังเกต: นิยามของ DFT ที่เราใช้ข้างต้นนั้นมี $\frac{1}{N}$ เป็น factor อยู่ด้วย ด้วยเหตุนี้จึงไม่มี $\frac{1}{N}$ factor ใน IDFT เราสามารถนิยาม

DFT ใหม่เพื่อการแปลงของ DFT และ IDFT คล้ายกันมากขึ้น ซึ่งทำได้โดยใช้ factor $\frac{1}{\sqrt{N}}$ ทั้งในสูตรของ DFT และ

IDFT นิยามทั้งสองแบบนี้เทียบเท่ากัน หากเราเลือกที่จะใช้นิยามใหม่ก็สามารถทำได้โดยการส่ง $\frac{1}{\sqrt{N}} s(t)$ แทน $s(t)$.

เมื่อเราใช้ IDFT แล้วก็ไม่มีความจำเป็นที่จะต้องมี oscillator สำหรับแต่ละคลื่นพาห่อย่อย เทคนิคการใช้ DFT/IDFT ร่วมกับ OFDM นี้เรียกว่า Discrete Multi-Tone (DMT) [Weinstein71] นอกจากนี้สิ่งที่ลดปัญหาความยุ่งยากของเครื่องรับส่งแบบ OFDM ก็คือ การนำเอา IFFT หรือ Inverse Fast Fourier Transform มาแทน IDFT โดย IFFT นั้นสามารถประมวลผลได้อย่างรวดเร็วเมื่อเทียบกับ IDFT ในทำนองเดียวกัน ส่วน DFT ของ OFDM ก็ถูกแทนที่โดย FFT (fast Fourier transform) การประมวลผลโดย FFT สามารถลดจำนวนการคูณในการแปลงแบบ DFT จาก N^2 เป็น $N \log N$ สำหรับข้อมูลที่มีขนาด N

นอกจากการนี้ยังมีการเพิ่มช่วงเวลาป้องกัน (guard time) เพื่อป้องกันผลกระทบที่เกิดจาก multipath fading ที่เรากล่าวถึงข้างต้น โดยมีการเสริมสัญญาณด้วย cyclic prefix ซึ่งเป็นการคัดลอกเอาสัญลักษณ์ส่วนท้ายของแต่ละ block ข้อมูลมาสอดไว้ก่อน block ข้อมูลนั้น

ปัญหาของ OFDM: อัตราส่วนกำลังงานสูงสุดต่อกำลังงานเฉลี่ย

ปัญหาที่สำคัญมากของระบบ OFDM ก็คือการที่สัญญาณมีอัตราส่วนกำลังงานสูงสุดต่อกำลังงานเฉลี่ย (Peak-to-Average Power Ratio, PAPR) ที่สูง เมื่อสัญญาณที่มีค่า PAPR สูงเข้าสู่วงจรขยาย (amplifier) ของเครื่องส่ง วงจรขยายจะต้องทำงานในช่วงอิมพัลส์และทำงานแบบไม่เป็นเชิงเส้น (non-linear) จึงเกิดการผิดเพี้ยนของสัญญาณ เป็นผลให้อัตราความผิดพลาดบิตข้อมูล (Bit Error Rate: BER) เพิ่มสูงขึ้น นอกจากนี้ยังทำให้ spectrum ของสัญญาณล้าออกนอกแถบความถี่ของช่องสัญญาณไปรบกวนช่องสัญญาณอื่น หากต้องการการขยายสัญญาณเชิงเส้น ก็จะมีผลต่อประสิทธิภาพทางกำลังของ amplifier ที่ใช้งาน ทำให้ต้องใช้ amplifier ที่มีราคาแพง

ในส่วนนี้เราจะใช้ทฤษฎีความน่าจะเป็นมาอธิบายว่าเหตุใดค่า PAPR จึงสูงตามค่า N ก่อนอื่น ถ้าเราย้อนกลับวิเคราะห์สัญญาณของ OFDM หลังจากผ่าน IFFT จะเห็นว่าสัญญาณมีค่าดังนี้

$$s[n] = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} s_k e^{+j\omega_k} \text{ โดย } \omega_k = 2\pi k \frac{n}{N} \text{ และ } n = 0, 1, \dots, N-1$$

ซึ่งโดยปกติแล้ว s_k จะเป็นจำนวนเชิงซ้อน หากมีการใช้คลื่นพาห่อย่อยจำนวนมาก (N มีค่าสูง) ก็มีโอกาสูงที่ $s[n]$ จะมีค่ามาก เพราะเป็นผลของการรวมกันของสัญญาณข้อมูลหลังการทำ IFFT ในที่นี้ถ้าเราสมมติให้ s_0, s_1, \dots, s_{N-1} เป็นตัวแปรสุ่มเชิงซ้อน (complex random variable) ซึ่งหมายความว่าแต่ละตัวแปรสุ่ม s_k จะสามารถแบ่งเป็นส่วนจริงคือ $\text{Re}\{s_k\}$ และ ส่วนเชิงซ้อนคือ $\text{Im}\{s_k\}$ โดยทั้งสองส่วนเชื่อมกันด้วยความสัมพันธ์

$$s_k = \text{Re}\{s_k\} + j \text{Im}\{s_k\}$$

ดังนั้นเราจึงมีตัวแปรสุ่ม $2N$ ตัวแปร ดังนี้

$$\text{Re}\{s_0\}, \text{Im}\{s_0\}, \text{Re}\{s_1\}, \text{Im}\{s_1\}, \dots, \text{Re}\{s_{N-1}\}, \text{Im}\{s_{N-1}\}$$

โดยตัวแปรสุ่มทั้งหมดเป็นจำนวนจริง

ในส่วนนี้เราจะให้ตัวแปรสุ่มทั้งหมดมีลักษณะเป็น iid (independent and identically distributed) โดย

$$\text{Re}\{s_k\} = \text{Im}\{s_k\} = 0 \text{ และ } \mathbb{E}\left[(\text{Re}\{s_k\})^2\right] = \mathbb{E}\left[(\text{Im}\{s_k\})^2\right] = \sigma^2$$

ซึ่งผลที่ตามมาก็คือ

- 1) $\mathbb{E}[s_k] = 0$
- 2) $\mathbb{E}[|s_k|^2] = \mathbb{E}\left[(\text{Re}\{s_k\})^2\right] + \mathbb{E}\left[(\text{Im}\{s_k\})^2\right] = 2\sigma^2$
- 3) $\mathbb{E}\left[(\text{Re}\{s_k\})(\text{Im}\{s_i\})\right] = 0$
- 4) ถ้าค่า $i \neq k$ แล้ว
 - a) $\mathbb{E}\left[(\text{Re}\{s_k\})(\text{Re}\{s_i\})\right] = 0$
 - b) $\mathbb{E}\left[(\text{Im}\{s_k\})(\text{Im}\{s_i\})\right] = 0$
 - c) $\mathbb{E}[s_i s_k^*]$ จะเป็น 0 เพราะ s_i กับ s_k นั้นเป็นอิสระ (independent) จากกัน ส่งผลให้

$$\mathbb{E}[s_i s_k^*] \stackrel{\text{(iid)}}{=} \mathbb{E}[s_i] \mathbb{E}[s_k^*] = 0 \times 0 = 0$$

ดังนั้นเราสามารถหาค่าคาดหมายของค่ากำลังงานเฉลี่ยได้เป็น

$$\mathbb{E}[|s[n]|^2] = \mathbb{E}[s[n] s^*[n]] = \frac{1}{N} \sum_{k=0}^{N-1} \sum_{i=0}^{N-1} \mathbb{E}[s_i s_k^*] e^{+j\omega_i} e^{-j\omega_k} = \mathbb{E}[|s_k|^2] = 2\sigma^2$$

นอกจากนี้เรายังสามารถกระจาย $s[n]$ ออกเป็นส่วนจริงและส่วนเชิงซ้อน โดยที่

$$\text{Re}\{s[n]\} = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} (\text{Re}\{s_k\} \cos(\omega_k) - \text{Im}\{s_k\} \sin(\omega_k))$$

$$\text{Im}\{s[n]\} = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} (\text{Re}\{s_k\} \sin(\omega_k) + \text{Im}\{s_k\} \cos(\omega_k))$$

หาก N นั้นมีค่ามาก เราสามารถใช้ทฤษฎีเซ็นทรัลลิมิตของ Lyapunov (Lyapunov Central Limit Theorem) ประมาณค่าความหนาแน่นเชิงความน่าจะเป็น (probability density) ของตัวแปรสุ่มทั้งสองด้วย ตัวแปรสุ่มแบบ Gaussian ที่มีค่าคาดหวังและค่าความเบี่ยงเบนเดียวกันกับตัวแปรสุ่มทั้งสอง ในที่นี้เรารู้แล้วว่า $\mathbb{E}[s_k] = 0$ สำหรับค่าความเบี่ยงเบนนั้น เพราะค่าคาดหวังเป็น 0 เราจึงสามารถหาค่าความเบี่ยงเบนจาก

$$\text{Var}[\text{Re}\{s[n]\}] = \mathbb{E}\left[\left(\text{Re}\{s[n]\}\right)^2 - \underbrace{\left(\mathbb{E}[\text{Re}\{s[n]\}]\right)^2}_0\right] = \mathbb{E}\left[\left(\text{Re}\{s[n]\}\right)^2\right]$$

ซึ่งถ้าเราแทนค่า $\text{Re}\{s[n]\}$ ด้วยสมการที่ประกอบด้วยพจน์ที่มี cosine และ sine ข้างต้น ค่าของ $\mathbb{E}\left[\left(\text{Re}\{s[n]\}\right)^2\right]$ จะเป็น

$$\mathbb{E}\left[\frac{1}{N} \sum_{k=0}^{N-1} \sum_{i=0}^{N-1} (\text{Re}\{s_i\} \cos(\omega_k) - \text{Im}\{s_i\} \sin(\omega_k)) (\text{Re}\{s_k\} \cos(\omega_k) - \text{Im}\{s_k\} \sin(\omega_k))\right]$$

หลังจากกระจายผลคูณออกมา เราสามารถกำจัดพจน์ที่มี $\text{Im}\{s_i\} \text{Re}\{s_k\}$ และ $\text{Re}\{s_i\} \text{Im}\{s_k\}$ ได้เพราะมีค่าคาดหวังเป็น 0 นอกจากนี้เมื่อ $i \neq k$ เรายังสามารถกำจัดพจน์ที่มี $\text{Re}\{s_i\} \text{Re}\{s_k\}$ และ $\text{Im}\{s_i\} \text{Im}\{s_k\}$ ได้ ดังนั้นสิ่งที่เหลืออยู่คือ

$$\mathbb{E}\left[\left(\text{Re}\{s[n]\}\right)^2\right] = \frac{1}{N} \sum_{k=0}^{N-1} \left(\mathbb{E}\left[\left(\text{Re}\{s_k\}\right)^2\right] \cos^2(\omega_k) + \mathbb{E}\left[\left(\text{Im}\{s_k\}\right)^2\right] \sin^2(\omega_k) \right)$$

เมื่อแทนค่า $\mathbb{E}\left[\left(\text{Re}\{s_k\}\right)^2\right] = \mathbb{E}\left[\left(\text{Im}\{s_k\}\right)^2\right] = \sigma^2$ เราจะได้

$$\mathbb{E}\left[\left(\text{Re}\{s[n]\}\right)^2\right] = \sigma^2$$

เราสามารถใช่วิธีการเดียวกันหาค่าอื่น ๆ ดังต่อไปนี้

- 1) $\mathbb{E}\left[\left(\text{Im}\{s_k\}\right)^2\right] = \sigma^2$
- 2) $\mathbb{E}\left[\left(\text{Re}\{s[n]\}\right)\left(\text{Im}\{s[m]\}\right)\right] = 0$
- 3) สำหรับ $n \neq m$ $\mathbb{E}\left[\left(\text{Re}\{s[n]\}\right)\left(\text{Re}\{s[m]\}\right)\right] = \mathbb{E}\left[\left(\text{Im}\{s[n]\}\right)\left(\text{Im}\{s[m]\}\right)\right] = 0$

ดังนั้นเราจึงสรุปได้ว่า เมื่อ N มีค่ามาก เราสามารถประมาณตัวแปรสุ่ม

$$\text{Re}\{s[0]\}, \text{Im}\{s[0]\}, \dots, \text{Re}\{s[N-1]\}, \text{Im}\{s[N-1]\}$$

ด้วยตัวแปรสุ่มแบบ Gaussian $\mathcal{N}(0, \sigma^2)$ ที่เป็นอิสระจากกัน ดังนั้นตัวแปรสุ่ม

$$|s[0]|^2, |s[1]|^2, \dots, |s[N-1]|^2$$

จึงมีลักษณะเป็น iid โดยมีการกระจายแบบ exponential เพราะ

$$|s[n]|^2 = \left(\text{Re}\{s[n]\}\right)^2 + \left(\text{Im}\{s[n]\}\right)^2$$

เป็นการรวมกันแบบ Euclidean ของตัวแปรสุ่ม Gaussian สองตัวแปรที่มีการแจกแจงแบบ iid โดยมีฟังก์ชันความหนาแน่น (pdf)

เป็นแบบเลขชี้กำลัง (exponential) ที่มีค่าคาดหวังเป็น $\mathbb{E}\left[|s[n]|^2\right] = 2\sigma^2$

เราจะนิยามค่า PAPR โดยสมการต่อไปนี้

$$\Gamma = \frac{\max_{0 \leq k \leq N-1} |s[n]|^2}{\mathbb{E}[|s[n]|^2]} = \frac{\max_{0 \leq k \leq N-1} |s[n]|^2}{2\sigma^2} = \max_{0 \leq k \leq N-1} \frac{|s[n]|^2}{2\sigma^2}$$

จะเห็นได้จากนิยามของ PAPR ว่าค่าเศษเป็นค่ากำลังงานสูงสุดและค่าส่วนเป็นค่ากำลังงานเฉลี่ย ดังที่ได้อธิบายไปแล้วข้างต้น เมื่อ N มีค่ามาก $|s[n]|^2$ จะมีการแจกแจงแบบ exponential เพราะฉะนั้น ตัวแปรสุ่ม $|s[n]|^2/2\sigma^2$ จึงมีการแจกแจงแบบ exponential ด้วย และมีค่าคาดหวังเป็น 1 เพราะ

$$\mathbb{E}\left[\frac{|s[n]|^2}{2\sigma^2}\right] = \frac{\mathbb{E}[|s[n]|^2]}{2\sigma^2} = \frac{2\sigma^2}{2\sigma^2} = 1$$

เมื่อเป็นเช่นนี้ ตัวแปรสุ่ม Γ จึงเป็นการหาค่าสูงสุดของตัวแปรสุ่มที่แจกแจงแบบ iid exponential จำนวน N ตัวแปร เราจึงสามารถหาความน่าจะเป็นที่ตัวแปรสุ่ม Γ จะมีค่าเกินจำนวนจริง γ ได้เป็น

$$P[\Gamma > \gamma] = 1 - (1 - e^{-\gamma})^N$$

เราจะสังเกตได้ว่าค่าความน่าจะเป็นนี้เพิ่มสูงขึ้นเมื่อ N ซึ่งก็คือจำนวนของคลื่นพหุข้อมีค่ามากขึ้น

มีงานวิจัยจำนวนมากที่นำเสนอวิธีการลด PAPR (PAPR reduction schemes) วิธีที่เราจะกล่าวถึงก็คือ Signal distortion technique ซึ่งใช้การบิดเบือนบริเวณส่วน peak ของสัญญาณแบบไม่เป็นเชิงเส้นตรง เพื่อไม่ให้สัญญาณมีค่ามากเกินไป โดยวิธีที่ง่ายที่สุดก็คือการ clip สัญญาณที่ระดับ γ ที่ตั้งไว้ก่อน [2, 3] ซึ่งเทียบได้กับการคั่นสัญญาณที่มีด้วย ฟังก์ชันรูปสี่เหลี่ยมซึ่งเราจะเรียกว่า rectangular window ที่มีความสูงเท่ากับ γ หากเราใช้วิธีนี้ แน่ใจว่าจะไม่มีสัญญาณที่มีขนาดสูงกว่า γ ออกมาจากเครื่องส่ง อย่างไรก็ตาม เมื่อสัญญาณถูก clip ไปก็จะมีรูปร่างไม่เหมือนเดิมซึ่งทำให้อัตราความผิดพลาดบิต (BER) ที่เครื่องรับสูงขึ้นไปด้วย นอกจากนี้ spectrum ของสัญญาณจะขยายขึ้น เพราะการคูณด้วย Rectangular window ในแกนเวลาเทียบเท่ากับการประสาน (convolve) สัญญาณด้วยฟังก์ชัน sinc เมื่อ spectrum ของสัญญาณขยายขึ้นสัญญาณก็จะไปรบกวนสัญญาณในย่านความถี่อื่น ดังนั้นเพื่อลดการไปรบกวนสัญญาณนอกแถบความถี่ จึงได้มีการเปลี่ยนเป็น window แบบอื่นที่มี spectrum ที่ดีกว่า เช่น การใช้ window แบบ Gaussian [7], cosince, Kaiser, หรือ Hamming วิธีการลด PAPR นั้นยังมีอีกหลายวิธี ตัวอย่างเช่น การใช้เทคนิค peak cancellation [8, 9] อีกวิธีหนึ่งก็คือการเข้ารหัสข้อมูลก่อนที่จะส่งสัญญาณเพื่อกำจัดกลุ่มสัญลักษณ์ที่มีค่า PAPR สูงเกินไป [1, 10]

ที่จริงแล้ว การพยายามลดค่า PAPR ให้ได้มากที่สุดนั้น อาจไม่ใช่สิ่งที่ถูกต้องนัก เพราะจุดมุ่งหมายที่แท้จริงของการสื่อสารก็คือการนำข้อมูลส่งไปถึงปลายทางอย่างถูกต้อง ดังนั้นนักวิจัยบางกลุ่มจึงให้ความสนใจกับการหา efficient PAPR ซึ่งก็คือ PAPR ที่ดีที่สุดภายใต้เกณฑ์ที่นักวิจัยสร้างขึ้น [11]

Reference

- [1] S. J. Shepherd, P. W. J. Van Eetvelt, C. W. Wyatt-Millington, and S. K. Barton, "Simple coding scheme to reduce peak factor in QPSK multicarrier modulation," Electron. Lett., vol. 31, no. 14, pp. 1131–1132, July 1995.

- [2] D. Wulich, "Peak factor in orthogonal multicarrier modulation with variable levels," *Electron. Lett.*, vol. 32, no. 20, pp. 1859–1860, Sept. 1996.
- [3] X. Li and L. J. Cimini Jr., "Effects of clipping and filtering on the performance of OFDM," *IEEE Commun. Lett.*, vol. 2, pp. 131–133, May 1998.
- [4] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 2nd ed. New York: McGraw-Hill, 1984.
- [5] D. Wulich, N. Dinur, and A. Gilinowiecki, "Level clipped high order OFDM," *IEEE Trans. Commun.*, vol. 48, pp. 928–930, June 2000.
- [6] Weinstein, S.B., Ebert P.M. *Data Transmission By Frequency Division Multiplexing Using the Discrete Fourier Transform.* *IEEE Trans. Commun.*, Oct 1971; COM-19; 5: 628-634.
- [7] Pauli, M., and H. P. Kuchenbecker, "Minimization of the Intermodulation Distortion of a Nonlinearly Amplified OFDM Signal," *Wireless Personal Communications*, Vol. 4, No. 1, January 1997, pp. 93–101.
- [8] De Wild, A., "The Peak-to-Average Power Ratio of OFDM," M.Sc. thesis, Delft University of Technology, Delft, the Netherlands, September 1997.
- [9] May, T., and H. Rohling, "Reducing the Peak-to-Average Power Ratio in OFDM Radio Transmission Systems," *Proceedings of IEEE VTC'98*, Ottawa, Canada, May 18–21, 1998, pp. 2774–2778.
- [10] Wilkinson, T. A., and A. E. Jones, "Minimization of the Peak-to-Mean Envelope Power Ratio of Multicarrier Transmission Schemes by Block Coding," *Proc. of IEEE Vehicular Technology Conference*, Chicago, IL, July 1995, pp. 825–829.
- [11] Wulich, D., "Definition of efficient PAPR in OFDM," *Communications Letters, IEEE* , vol.9, no.9, pp. 832-834, Sep 2005